A Freight Stop Purpose Model Using Enriched GPS Data

Monique Stinson, PhD, Corresponding Author
Office of Statistical and Economic Analysis
Bureau of Transportation Statistics
Office of the Secretary of Transportation (OST-R)
United States Department of Transportation
1200 New Jersey Avenue, SE
Washington, DC 20590, USA
Email: monique.stinson@dot.gov

Olcay Sahin, PhD
Vehicle and Mobility Systems
Argonne National Laboratory
9700 South Cass Avenue
Lemont, Illinois 60439, USA
Email: osahin@anl.gov

ORCID:
Monique Stinson: 0000-0003-1337-1903
Olcay Sahin: 0000-0003-4235-036X

Presented at the 2024 Transportation Research Board Annual Meeting

ABSTRACT

Truck tour data are important for understanding freight truck operations and their impacts on the economy, congestion, and sustainability. Tour data capture the movements of trucks as they travel from their depots to make various stops before returning to their depots. Understanding the purpose of these stops (pickup, delivery, rest, etc.) is an essential component of freight truck tour analysis. Evaluating stop purposes is especially critical to evaluate truck electrification, resilience, supply chains, truck parking, and other emerging topics. Many studies use truck GPS data to infer truck activity, but most analyze trips instead of tours. Also, GPS data generally don't have stop purpose information. Consequently, truck tour data with stop purpose information are a major freight data gap. We address this gap by developing a process to infer both truck tours and stop purposes. Our main contribution is developing a behavioral model that allows analysts to infer stop purposes and tour patterns. To do so, we first fuse truck GPS data with geospatial data on businesses and interstates. We enrich the new dataset by manually labeling the estimated stop purposes. We use the enriched data to develop behavior models that predict the purpose of each stop. Analysts can apply the resulting model to predict stop purposes for other GPS data, thereby filling an important need in truck touring data development.

Keywords: stop purpose, truck tour, discrete choice, econometric, behavior, GPS, telematics, data fusion, depot

1. INTRODUCTION

This study focuses on developing data on freight truck tours with stop purpose information. Similar to earlier works including [1] and [2], we define a tour as a truck journey that begins and ends at a depot (or home base) while making one or more stops before returning to the depot. Tour data provide comprehensive information on truck activities throughout the day, including information on stop locations, stop activities, stop dwell times, routes used, and time spent at the depot. The ideal tour dataset also contains information on commodities transported and pickup/delivery locations, but these are out of scope to this study. Truck tour data are essential for applications such as truck touring models [3] and truck activity-based models, which use a daily activity pattern [4].

Sources of freight truck tour data include truck driver surveys and fleet telematics data with stop addresses and commodity information. However, since companies are usually reluctant to share this information, many previous efforts cover a limited number of few fleets (e.g., [5], [6]) or are limited to one area [7]. Alternatively, large-scale GPS data are potentially a rich source of tour data, especially when the data cover numerous fleets, a large geographical area, and multiple days of operations for each truck. GPS data that meet these criteria have become readily available in recent years. For these reasons, using GPS data to study truck tours is very desirable. However, without supplemental data, it is difficult to infer depot locations and stop purposes from GPS data alone. Therefore, truck touring data with stop purpose information remains a major freight data gap.

Tour data are important for freight analysis because they provide a detailed picture of truck operations. Such data are necessary for many studies including urban freight operations [8], drayage truck emissions [9] and truck electrification [10], and regional network simulations [3]. For example, truck electrification is beginning primarily with charging stations at depots, with few publicly available charging stations elsewhere. Consequently, tour data are essential for evaluating whether trucks can complete their daily activities between charges before the battery is depleted.

Tour data with information on stop purposes and depot locations is difficult to develop. First, the depot for each truck must be identified. This is needed to construct a truck tour dataset that can be further analyzed. When all data come from one fleet with known depot locations, this step is trivial. However, many data sources include multiple fleets with unknown depot locations. Second, the purpose of other stops must be inferred. Stop purpose is difficult to describe in detail. For example, at a “rest stop”, a driver may be resting, eating, or refueling. Therefore, in this study, we focus on identifying the type of establishment that the trucks visit: rest, retail, manufacturing, intermodal, trucking, warehouse or distribution center (WDC), and other.

Unfortunately, neither depot location nor stop purpose can be inferred from GPS data alone. Previous studies devised solutions to developing truck tour data with depot and stop purpose information. The data gap has been addressed by supplementing GPS data with land use and establishment data [11] or carrier logs (e.g., [6], [12]). Even so, without manual intervention, stop purpose inference is limited [13]. More importantly, methods to identify depots and stop purposes are still a gap. The Literature Review discusses previous studies in more detail. In brief, existing methods provide valuable insights into truck activity using GPS-based data. However, they have some limitations, such as focusing only on modeling truck industry type (e.g., [13] and [11]) or focusing only on drayage truck fleets [9].

The objective of this work is two-fold. First, we fuse numerous data sources together to supplement GPS data for purposes of analyzing truck activity. Second, we develop a stop purpose model that can be used to evaluate stop purposes in a tour. The resulting model can be applied in conjunction with our depot detection model (see Literature Review) to construct a truck tour dataset that contains depot locations and stop purposes for individual truck tours.

Our approach involves the following elements. First, we develop a rich dataset of truck travel by fusing GPS data with geospatial datasets of business establishments and the US interstate highway system. We label stops using judgment in combination with satellite imagery. We analyze the fused data and present descriptive statistics to characterize trucks and their stops. Second, we use the fused data to estimate a series of multinomial logit (MNL) models [14] that predicts the purpose of each stop. Unlike earlier works, we use the entirety of our passive input sources to predict stop purposes: GPS-based truck operational attributes, establishment locations and NAICS (North American Industry Classification System) data, interstate data, and a list of names of popular truck fuel and rest stops. Moreover, we estimate six MNL models in total. In each version, we introduce additional variables into the model. This provides insight into what types of variables are the strongest predictors of stop purpose.

This work has the following limitations. First, the core dataset is limited to a relatively small sample of 219 trucks and the 4,852 unique stops that they make. Second, we lack information about what types of truck fleets are included in the sample. We do not know if the fleets are mostly for-hire or private fleets, whether they are mostly large/small fleets, and so on. Third, we apply MNL models, but since the data are panel data, the MNL model may have bias since it does not account for panel effects. We plan to examine this effect using a mixed logit model in follow-up efforts.

The remainder of this paper is structured as follows. First, we describe previous literature and how our study differs. Second, our data sources are presented. Third, the methodology is outlined. Fourth, the MNL stop purpose model findings are presented and discussed. We conclude the paper by summarizing key results and discussing extensions and limitations.

2. LITERATURE REVIEW

This section discusses earlier truck GPS studies. The discussion focuses mostly on truck GPS studies that infer or analyze stop purposes. We then position our study in the context of these earlier works and to highlight the unique contributions of our study.

Truck GPS data can often be used to extract tour information if operations over the course of a day or weeks can be extracted for individual trucks. Truck GPS data usually consist of a truck ID, a series of points (typically xy coordinates) that the truck has traversed, and timestamps, or time and date that each point is observed. Additional insights, such as routes used or speed and acceleration profiles, can also be extracted. As such, GPS data represent the movement of a set of trucks throughout the day or week.

Extracting tours from GPS data typically involves the following steps. First, a stop detection algorithm (such as DBSCAN [15]) is applied to infer where stops occur. Second, the analyst must identify which stop is the home base, or depot, of each truck. Third, the data are transformed into tour data. Each tour is labeled with a tour ID, depot location, time and location of intermediate stops, and potentially other information.

GPS data have been widely used to evaluate truck activities. However, due to the difficulty of inferring tours, most GPS-based analysis focuses on trips instead of tours. Each trip, which is a journey from an origin to a destination, is one leg of a tour. Trip data are sufficient for evaluating many phenomena, including truck traffic patterns and system performance such as travel time reliability [16]. Truck GPS data are widely used for truck trip data analysis, as methods for extracting stop locations and trips from GPS data are well developed and readily available [17]. GPS data are also used to differentiate between pickup/delivery and rest/refueling stops (e.g., [18]).

Identifying the depots that are embedded in truck GPS data is much more challenging. Previous efforts have employed a range of assumptions to help identify depots. [6] and [19] assume that the depot is the most frequently visited location. However, our analysis shows that sometimes the most frequently visited location is not a depot—e.g., for some trucks it is a particular rest stop. [9] treat the largest cluster at the end of the day as the depot. However, their analysis focused on drayage trucks and may not be applicable to GPS data more generally, which can include long-haul trucks with geographically fragmented stop locations. [20] found that depots are associated with lower entropy, or fewer unique trucks, in comparison with rest stops.

To address these limitations, we developed a logistic regression model to identify depots [21] to remedy some of these issues. This work was presented at an earlier conference; therefore, we present it as a previous study and discuss it here in the Literature Review section. We first developed an enriched dataset of truck tour-related variables (see Data section), then used manual supervision to evaluate which stop is the depot of each truck. Finally, we used logistics regression to create a depot detection model. The resulting model (**TABLE 1**) can be applied to predict the depot of a given truck using the enriched dataset as the input. McFadden's pseudo-r2 measure yields a goodness of fit of 0.353 based on 4,852 stop observations, which cover the activities and depots of 219 unique trucks. The results indicate that the number of visits and cumulative time spent by the truck are important predictors for identifying the depot. The number of visits made by other trucks is statistically insignificant in this model. When the truck stop is located near a likely fuel/truck rest establishment and is near the interstate, the stop is unlikely to be a depot. Concentrations of manufacturing, trucking, mining, utilities, construction were all associated with positive depot identification, whereas stop locations in retail-heavy areas are less likely to be depots.

TABLE 1 Depot Detection Model [21]

Parameter	Estimate	*Std. Error*	z value	Pr(>|z|)
Intercept	-2.687	*0.211*	-12.71	***
Number of visits by…				
…this truck	0.132	*0.016*	8.08	***
…other trucks	0.0005	*0.0007*	0.76	
Stop is near a major fuel/truck rest establishment	-0.336	*0.195*	-1.72	.
Cumulative duration (hours) spent by truck at this stop	0.027	*0.002*	12.21	***
Number of establishments at this stop by NAICS code:				
NAICS 2	0.120	*0.053*	2.25	*
NAICS 3	0.017	*0.008*	2.25	*
NAICS 44-45	-0.031	*0.008*	-3.69	***
NAICS 484	0.031	*0.014*	2.21	*
Stops per day by this truck	-0.238	*0.037*	-6.41	***

Furthermore, information on stop purposes is very useful for analyzing truck tours. However, methods to identify stop purposes using GPS data remain a gap. The most relevant prior studies fused GPS data with satellite imagery [13] and Geographic Information systems, or GIS, datasets [11]. These works infer what industry is served by the truck. Although these works provide important insights, their main objective is using the GPS data to characterize the industry that the truck serves, rather than identifying depots and other stop purposes as a part of truck tour data development. Moreover, the algorithm of [13] uses explanatory variables based only on the truck GPS data. The algorithm produces very good predictions of industry type, supporting the conclusion that various GPS-based truck operational attributes (number of stops made, stop duration, etc.) are good predictors of truck industry type. However, it is not clear whether the methods of these studies would yield the same level of accuracy to predict the purpose of individual stops.

In another work, [22] fuses truck GPS with point-of-interest (POI) data to distinguish pick-up/delivery from rest/refueling stops. While relevant to the current study, this work does not supply the level of stop purpose detail that we seek. In conclusion, truck touring data with detailed stop purpose information remains a significant gap despite its importance to freight analysis. We contribute to the literature by developing a discrete choice model that predicts stop purposes of freight trucks using GPS data and other sources.

3. DATA

3.1. Data Sources

The primary data source for our study is a truck GPS data sample obtained from the American Transportation Research Institute (ATRI). The sample has four parts. Two parts are based on a random sample of US trucks. The other two parts are a sample of all trucks in the ATRI database that operate in the Chicago Metropolitan Agency for Planning (CMAP) region. Data were obtained for three-week periods in March and August for years 2014 and 2017. In total, the sample contains 462,156 unique trucks and 231,854,183 GPS pings.

We enrich the GPS data with a geospatial layer of US interstate highways [23]. We further integrate data from numerous establishment-level datasets, which contain business names, geospatial data and NAICS information: CoStar [24], Yahoo! Finance [25], Bloomberg [26], FleetSeek [27], and Dun and Bradstreet [28]. Finally, we create a set of popular truck fuel and rest station names, including:

- 7-Eleven, Amoco, Arogas, BP, Casey, Cenex, Chevron, Circle K, Citgo, Conoco, Exxon, Family Express, Flying J, Iowa 80, Koa, Kum & Go, Kwik Stop, Love's, Marathon, Maverik, Meijer, Mobil, Phillips 66, Pilot, Quik Stop, Ranger, Shell, Sinclair, Speedway, Sunoco, TA (Travel America), Texaco, Thornton, Valero, United Refining
- Names that ended in "Mart" or had the words "Plaza", "Convenience", "Energy", "Fuel", "Hiway", "Pantry", "Petro", "Rest Stop", "Travel", "Truck Stop", "Visitor"

The final data source was satellite imagery in Google Maps [29]. Specifically, we used the aerial images of buildings, roadway and parking networks, and place names to verify stop purposes and depots.

3.2. Initial Data Processing

We use the DBSCAN algorithm [15] to detect stops that are embedded in the raw GPS data. Algorithm parameters of 300 seconds (five minutes) with a 500-meter buffer were used. This means that if a truck stays within a 500-meter buffer over a five-minute period, then the cluster of points is assumed to be a stop. The resulting dataset contains 3,390,722 stops. It also contains stop characteristics: stop coordinates (mean lat/long), duration (seconds), start time and end time.

We then derive other truck and stop characteristics from the stops dataset. First, we create a 500-meter buffer around each stop, then count the total number of other trucks that make a stop in the buffer. This gives us a measure of entropy [20] for each stop. Second, we derive a value that we call "travel diameter" as the maximum Great Circle Distance between any two stops in the dataset. Third, for each truck, we compute the total vehicle-miles traveled (VMT), the average daily VMT, total number of stops, average daily number of stops, and average distance between stops.

We then collected business establishments data from multiple sources as described above. We merged all establishments data into a single dataset, then removed duplicate records. For example, if a Love's truck stop at 123 Elm Street was present in two or more sources, then only one record of this establishment was kept.

3.3. Data Fusion

This subsection discusses the various ways that the stops data were fused with other data sources. The goal of this is to create a rich dataset that can be used for two things. First, the enriched data will be used to estimate the stop purposes model. Also, as the Literature Review describes, the enriched data were used to estimate the depot detection model. Second, the enriched data can be used to estimate stop purposes for out-of-sample stops.

First, stop data were fused with the establishments data. This involved creating an 800-meter buffer around each stop, then counting the total number of establishments and the number by NAICS code within the buffer. Since the buffer area is constant, the number of establishments is effectively the same as the density of establishments per square kilometers. Our model uses the former for simplicity.

Second, the stops data were analyzed jointly with the interstates shapefile. For each stop, the distance between the stop and the nearest interstate was computed.

Third, the establishments and interstates data were combined and processed to form a subset of establishments that (a) have a popular truck fuel/rest stop name and (b) are located within 1.2 km of an interstate. The establishments in this subset are treated as "likely" fuel/rest stops and are used accordingly as an input to the stop purpose model.

Even with this greatly enriched dataset, it is difficult to verify the stop purpose, or which establishment the truck visits. There are several reasons for this. First, trucks often park some distance away from the establishment. This can happen, for instance, when the parking lot is already full. As a result, the nearest establishment is not always the one that was visited. Also, truck parking (or loading / unloading bays) can be located at the rear of the establishment, which may be close to other freight-related establishments. Third, establishments can be closely clustered together or even may reside in the same building.

For these reasons, we need to use additional "ground truth" data to estimate the stop purpose model with confidence. We developed a process to further enrich the data using satellite imagery. Using the imagery as context, a human analyst adds additional detail to each truck stop record. This process is discussed next.

3.4. Manual Labeling

We randomly selected a subsample of 369 trucks from the full dataset, then manually verified the stop purpose and depot for each of these trucks. We later use the resulting dataset for in-depth analysis and model estimation. To facilitate the manual verification process, we developed a visualization and database development tool (**Figure 1**) using RShiny [30]. Two analysts manually labeled each stop for these 369 trucks using the process shown in **Figure 2**. Information from the stops and establishments databases and aerial imagery are read and visualized by the RShiny viewer. The human analyst examines each stop that was detected, then labels each stop based on land use according to her/his judgment. Thirteen land uses were provided to identify land use: Trucking Facility (Depot or Terminal), Intermodal, Maintenance / Inspection, Manufacturing (Manuf.), combined Manufacturing/WDC site (Manuf./WDC), Multi-purpose Facility, Not a Stop, Other Freight Stop, Residential, Rest, Retail, WDC/Storage-1 company, WDC/Storage-Multi-tenant, Missing or Unknown.

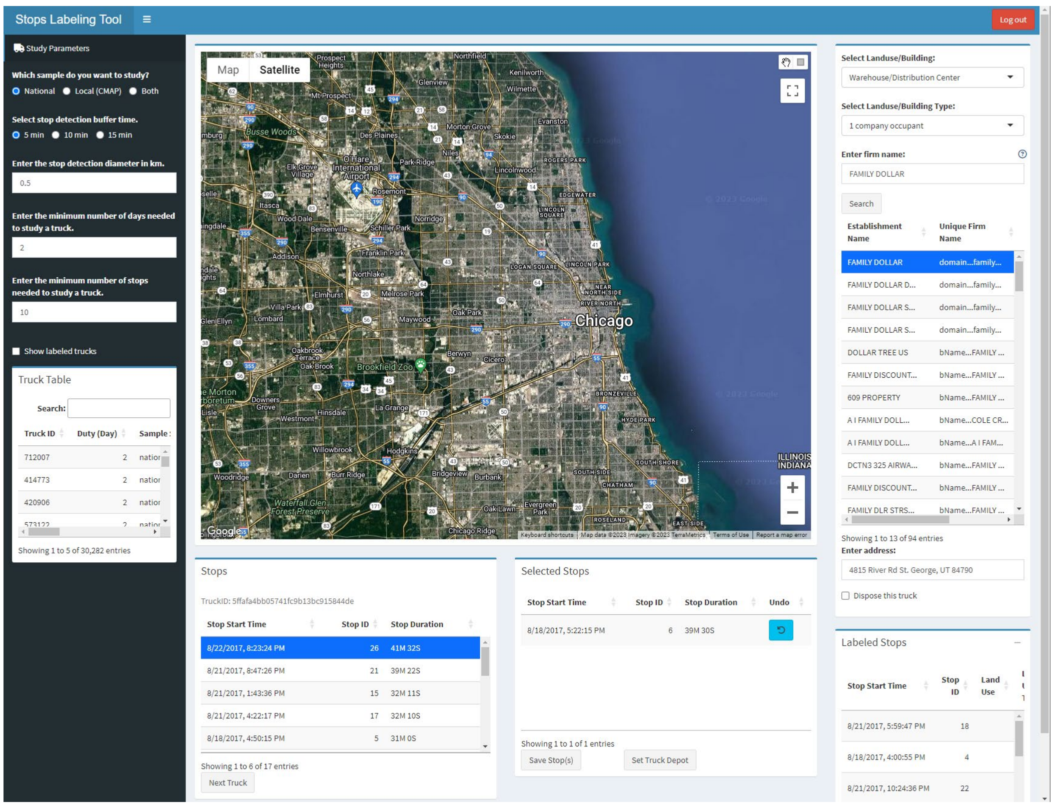


Figure 1 RShiny viewer for labeling stop purposes and identifying depots

After tagging all stops that a truck makes, the analyst was asked to infer which stop is the truck's depot. In some cases, a depot could not be identified. These trucks were removed from the analysis. A total of 219 trucks and their 4,852 stops remain. We use these observations later for model estimation. As a last step, the user also confirms whether clusters of stops are associated with the same place or with different places. Future efforts can instead use clustering algorithms to join stops together into a single cluster.

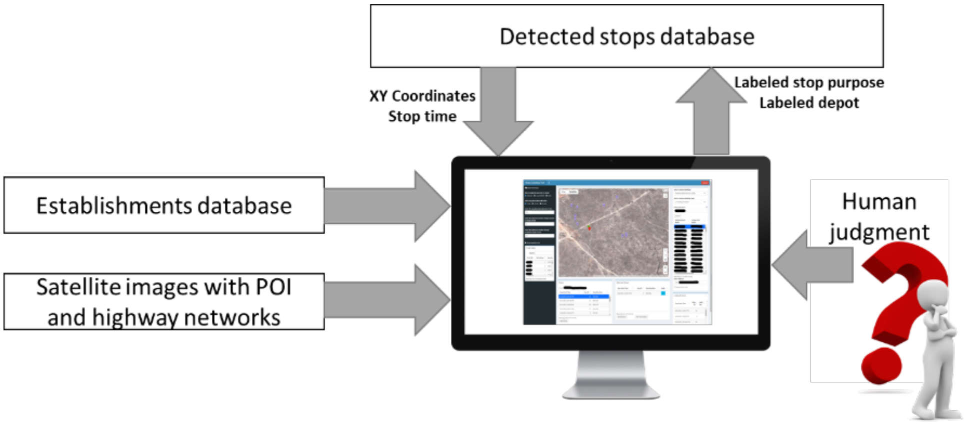


Figure 2 RShiny input/output and human labeling process

TABLE 2 shows the number of stops labeled for each stop purpose. Most stops are associated with a freight-related land use, with 35 percent located at a storage or distribution facility (WDC), 13.5 percent at a retail location. 10 percent at a trucking facility (depot or terminal), about 7.5 percent at a manufacturing facility, and five percent at an intermodal facility. A large percentage (22%) are at a rest or fuel station, and nearly five percent of the detected stops actually were not a stop. The last situation can happen when, for example, a truck is stuck in traffic.

TABLE 2 Number of stops by stop purpose (land use type)

Land Use (Stop Purpose) Type	Total Stops by Tagged Trucks	Total Stops by All Trucks	Percentage of Tagged Stops
Depot or Terminal	1,062	24,687	10.86
Intermodal	443	20,659	4.53
Maintenance/Inspection	39	789	0.40
Manuf.	93	1,520	0.95
Manuf./WDC	624	8,444	6.38
Multi-purpose Facility	6	75	0.06
Not a Stop	451	642	4.61
Other Freight Stop	82	334	0.84
Residential	39	81	0.40
Rest	2,169	241,290	22.18
Retail	1,324	12,203	13.54
WDC/Storage-1 company	2,745	77,137	28.06
WDC/Storage-Multi-tenant	704	17,105	7.20

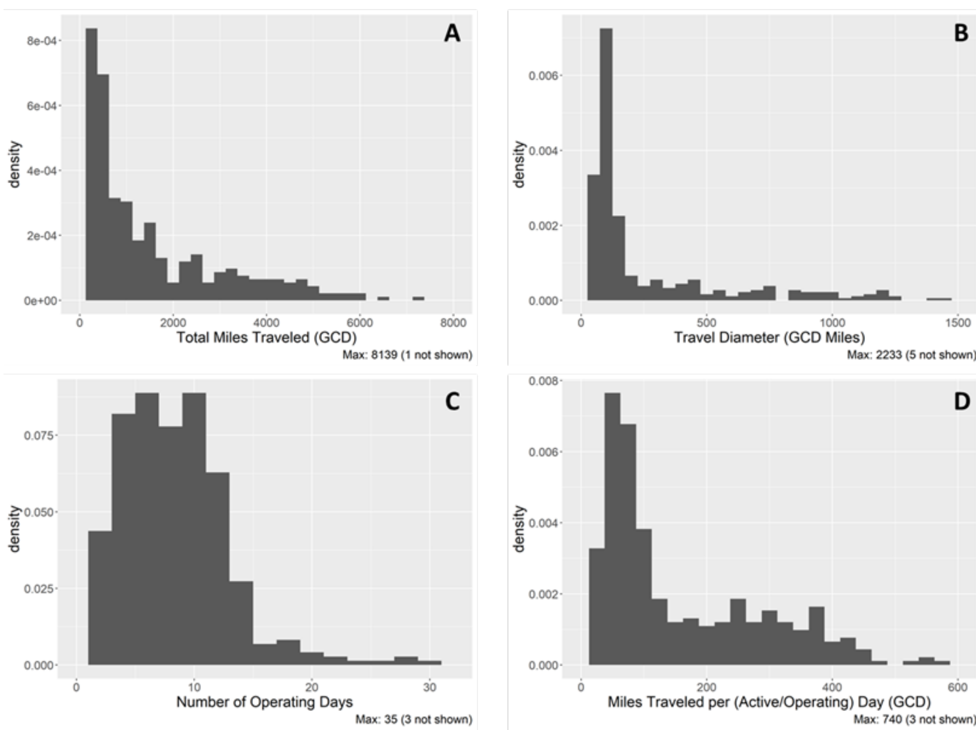
*Manuf. = manufacturing
WDC = warehouse or distribution center

TABLE 2 also shows the number of stops detected by all trucks within a 500-meter buffer of the stops made by the 219 trucks. As expected, some types of detected stops (Not a Stop and Residential, for instance) do not have many "visits" by other trucks. In contrast, stops at freight-related facilities (e.g., Intermodal or Manufacturing) and rest or refueling locations are visited by thousands of other trucks from the entire sample.

3.5. Descriptive Statistics

The next several figures and table provide descriptive statistics, showing the distribution of various operational features across the 219-truck sample. **Figure 3** shows the distribution of miles traveled using straight line, or Great Circle Distance (GCD) in the focused sample, including total miles traveled for the duration of the sampling period and average VMT on each day of operation. Panel D indicates that many of the sampled trucks provide local or regional service, traveling less than 100 to 200 miles per day, and that long-haul trucks are also well represented in the sample. Similarly, most trucks in the sample have a travel diameter of 200 miles or less, indicating local service, with a fair number of regional or long-haul service represented in the sample.

Figure 3 Operational characteristics: miles traveled and days in operation



Note: GCD (Great Circle Distance): straight-line distance that accounts for Earth's curvature

Figure 4 shows the following distributions: the total number of stops detected, stops per mile, and stops per day of operation for each truck. Panel B shows that a large percentage of trucks have less than 0.05 stops per mile, or less than one stop for every 20 miles (of straight-line distance). Panel C shows that most trucks in our sample have between one and five stops per day of operation, while some trucks have 15 or even more.

Figure 4 Operational characteristics: Number of stops made

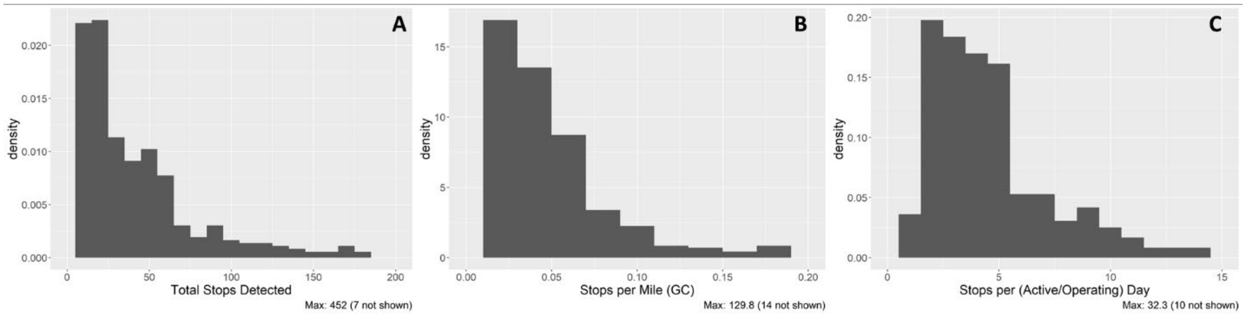


Figure 5 labels Manufacturing and Retail stops as Freight Production/Consumption (Prod./Cons.) and stops with a logistics-related function such as intermodal transferring or distribution (Freight Logistics). As this figure shows, rest stops, freight logistics stops, and maintenance or inspections stops tend to attract large numbers of unique trucks. Conversely, stops with a Freight Production/Consumption business activity tend to attract fewer unique trucks. Furthermore, stops with a residential or maintenance/inspection purpose tend to have a relatively long duration. Other stops are shorter on average, with some having a very long tail (especially rest stops and Freight Logistics stops, which also contains truck depots and terminals). Some “Not a Stop” observations appear to be very long. Upon further investigation, we found that these stops happen at the regional boundaries (e.g., the US/Canadian border), where GPS tracking ceases. These records are eliminated from the model estimation.

Figure 5 Unique trucks and stop duration boxplots by stop purpose

TABLE 3 summarizes several stop attributes over the 219-truck sample. Average stop duration is about 2.5 hours, but stop duration overall ranges from a few minutes to many days. The typical stop is about four miles from the nearest interstate. About 83 unique trucks on average are detected at a given stop. The sampled trucks tend to stop in low-density areas (where density is based on establishment density).

TABLE 3 Additional stop characteristics

Variable	5th percentile	25th percentile	Mean	75th percentile	95th percentile	StdDev
Stop Duration (hours)	0.11	0.27	2.48	1.14	12.40	7.64
Distance to interstate (miles)	0.02	0.16	4.11	2.36	21.35	12.47
Number of unique trucks detected at stop	1.00	7.00	83.46	110.00	336.00	122.16
Land use density (establishments per sq. mi.)	0.00	10.13	64.81	93.67	212.66	79.11
Number of NAICS 1-2 establishments within 800 meters	0.00	0.00	1.13	1.00	5.00	1.98
Number of manufacturing (NAICS 3) establishments within 800 meters	0.00	1.00	10.10	13.00	41.00	14.97
Number of wholesale (NAICS 42) establishments within 800 meters	0.00	0.00	2.10	2.00	10.00	4.13
Number of non-store retailers (NAICS 454) establishments within 800 meters	0.00	0.00	0.05	0.00	0.00	0.24
Number of other retail (NAICS 44-45) establishments within 800 meters	0.00	2.00	16.41	21.00	59.45	24.38
Number of trucking establishments (NAICS 484) within 800 meters	0.00	0.00	3.22	4.00	13.00	5.11
Number of warehousing and storage (NAICS 493) establishments within 800 meters	0.00	0.00	4.31	5.00	19.00	7.84
Number of other transportation and warehousing (Other NAICS 48-49) establishments within 800 meters	0.00	0.00	2.44	3.00	9.00	4.93
Number of other (NAICS 5+) establishments within 800 meters	0.00	1.00	11.44	16.00	39.00	16.84

4. METHODOLOGY

The estimation dataset comprises 219 trucks and the 4,852 unique stops that they make. Our stop purpose model uses the widely known Multinomial Logit model [14]. Our MNL model, which is based on Random Utility theory, uses the following utility equation (**Equation 1**):

$$U_{i,t}(Stop\ Purpose\ P) = \alpha_P + \beta_{1,P} * StopStatistics_{i,t} + \beta_{2,P} * OperationalAttributes_{i,t} + \gamma_P * InterstateProximity_i + \delta_P * EstablishmentStatistics_i + \eta_P * LikelyFuelRest_i + \epsilon_{i,t} \quad (1)$$

where:

$U_{i,t}(Stop\ Purpose\ P)$ is the utility of stop purpose P for individual stop i that truck t makes,

$StopStatistics_t$ is a vector of stop statistics (number of unique trucks visiting stop i where truck t also stops; cumulative hours spent by a given truck t at stop i),

$OperationalAttributes_{i,t}$ is a vector of truck operational attributes (miles per day; stops per day; and travel diameter) associated with truck t stopping at stop i,

$InterstateProximity_i$ is the scalar distance between the stop and the interstate,

$EstablishmentStatistics_i$ is a vector containing statistics on establishments in the specified buffer around the stop (total number of establishments; number of retail establishments with NAICS 44-45; number of trucking establishments with NAICS 484; number of warehouse/distribution establishments with NAICS 493),

$LikelyFuelRest_i$ is a scalar binary variable that equals one if the stop is considered a likely fuel or rest stop and is zero otherwise,

$\alpha_P, \beta_{1,P}, \beta_{2,P}, \gamma_P, \delta_P, and\ \eta_P$ are purpose-specific parameters to be estimated, and

$\epsilon_{i,t}$ is an error term that is independently and identically distributed across observations and that includes unobserved factors and measurement error.

The MNL model assumes that the utility is linear in parameters. It is estimated by forming a probability expression that applies to each observation (**Equation 2**):

$$P_i(Stop\ Purpose\ P) = \frac{\exp(\psi' x_{iP})}{\sum_{j\epsilon C_P} \exp(\psi' x_{jP})} \quad (2)$$

where ψ is a vector of all parameters to be estimated, x is a vector of all explanatory variables, and C_P is the choice set consisting of all stop purposes.

We use Biogeme [31] to estimate the model parameters. The parameters are estimated by forming the log likelihood, then using maximum likelihood estimation. For more details, the reader can refer to [31] or [14].

5. FINDINGS

We now present and discuss the model estimation results. First we show the results from the incremental estimation process, where we added a new set of variables in each round. We then show the results of the best model.

The model is estimated incrementally. A new group of variables is added in each iteration. **TABLE 4** shows the resulting impacts on log likelihood and the estimated goodness of fit using the standard $\bar{\rho}^2$. The constants-only model (Model #1) has a goodness of fit of 0.211. Introducing stop-related variables ($StopStatistics$ in #2) improves the fit by about 0.128 to 0.339. Introducing the truck operational variables (#3) further improves fit by 0.031 to 0.370. Adding the distance to interstate variable (#4) has a negligible effect on model fit. However, adding the fused stop-establishment information (#5) yields another notable model improvement (fit improves by 0.050), while adding the lone "likely fuel/rest stop" indicator variable generates an additional modest but noticeable improvement (fit improvement of 0.008).

TABLE 4 Results of the Incremental Model Estimation

Explanatory Variables to Include	Model Version					
	1	2	3	4	5	6
Constants	X	X	X	X	X	X
Stop-related variables (source: GPS data)		X	X	X	X	X
Truck operations variables (source: GPS data)			X	X	X	X
Fused stop / interstate variables				X	X	X
Fused stop / establishments variables					X	X
Fused stop / establishments / interstate variables						X
Final log likelihood* (*Null log likelihood = -10089.45)	-7956.968	-6650.788	-6316.548	-6299.195	-5776.646	-5694.447
Likelihood ratio test	4264.964	6877.325	7545.805	7580.511	8625.609	8790.008
Goodness of fit ($\bar{\rho}^2$)	0.211	0.339	0.370	0.371	0.421	0.429

TABLE 5 shows the preferred model after eliminating very insignificant variables and combining explanatory variables that have similar effects. Robust standard errors, t-tests, and p-values are provided. We estimate purpose-specific parameters to gauge the different effect of each variable on predicting each stop type. In other words, the second column of the table indicates which utility equation the parameter belongs to. For instance, the "Likely Fuel / Rest Stop" parameter is only included in the Rest Purpose utility equation. We do not discuss the results in detail given space limitations. However, we note that the estimated parameter values and signs are generally intuitive. The resulting estimates can be applied to predict stop purposes for out-of-sample stops based using only the unlabeled explanatory data.

6. CONCLUSIONS AND LIMITATIONS

This paper presents a stop purpose model that was developed using an enriched truck GPS dataset. The model performs well even though the sample size for our calibration dataset is fairly limited, with 219 trucks and about 5,000 stops. Our results demonstrate that most of the enrichment yields substantial improvements to model fit, and that the most impactful variables are stop-related variables derived directly from the GPS data.

The next step of this work is to improve the theoretical aspects of the model by estimating its mixed logit variant. The input dataset contains two types of panel effects: one is due to repeated observations by unique trucks, and another is due to repeated observations on the same tour. Other improvements involve enriching the data with more publicly available datasets, such as the Bureau of Transportation Statistics National Transportation Atlas Database layers [32]. Increasing the sample size is another potential extension. Finally, introducing information that will enable improved commodity type inference and developing a model or algorithm to infer commodity type will also benefit studies that seek to leverage passive GPS data sources for trucking analysis.

TABLE 5 Preferred Model

Parameter Name	Description	Estimate	Rob. Std Err.	Rob. t-test	Rob. p-value
	Alternative Specific Constants				
ASC_IMX	Intermodal	0.39	0.374	1.05	0.295
ASC_MANU	Manuf.	0.29	0.197	1.47	0.142
ASC_NS	Not a Stop	0.12	0.185	0.654	0.513
ASC_REST	Rest	1.80	0.207	8.66	0
ASC_RET	Retail	1.55	0.224	6.9	5.08E-12
ASC_TRUCKING	Trucking (terminal or depot)	0.31	0.263	1.16	0.245
ASC_WDC	WDC	1.91	0.179	10.6	0
	Total Number of Unique Trucks Detected at Stop				
B_AuniqTrks_NonRet	Non-Retail	0.032	0.010	3.26	0.00111
B_AuniqTrks_REST	Rest	0.036	0.010	3.66	0.000252
B_AuniqTrks_RET	Retail	0.011	0.010	1.07	0.285
	Cumulative Hours Spent at Stop (by This Truck)				
B_CumHrs_IMX	Intermodal	-0.0091	0.0064	-1.42	0.156
B_CumHrs_MAN	Manuf.	-0.0113	0.0042	-2.69	0.00709
B_CumHrs_NS	Not a Stop	-0.0149	0.0058	-2.55	0.0107
B_CumHrs_REST	Rest	-0.0213	0.0040	-5.28	1.32E-07
B_CumHrs_RET	Retail	-0.0311	0.0107	-2.91	0.00358
B_CumHrs_TRK	Trucking (terminal or depot)	0.0038	0.0022	1.71	0.0871
B_CumHrs_WDC	WDC	-0.0056	0.0025	-2.21	0.0274
	Stops per Day (by This Truck)				
B_StopsPerDay_IMX	Intermodal	-0.08	0.0325	-2.38	0.0175
B_StopsPerDay_REST	Rest	-0.11	0.0163	-6.99	2.78E-12
B_StopsPerDay_RET	Retail	0.04	0.0142	2.57	0.0102
B_StopsPerDay_TRK	Trucking (terminal or depot)	-0.13	0.0358	-3.53	0.000417
	Travel Diameter (miles) (of This Truck)				
B_TravDiam_IMX	Intermodal	-0.0059	0.0019	-3.13	0.00175
B_TravDiam_NS	Not a Stop	0.00022	0.00014	1.5	0.133
B_TravDiam_REST	Rest	0.00047	0.00011	4.11	4.04E-05
B_TravDiam_RET	Retail	-0.0012	0.0002	-6.12	9.09E-10
B_TravDiam_WDC	WDC	-0.00062	0.0001	-4.61	4.11E-06
	Number of Manufacturing (NAICS 31-33) Establishments in Buffer				
D_n3_Estabs_IMX	Intermodal	0.043	0.0168	2.56	0.0104
D_n3_Estabs_MAN	Manuf.	0.109	0.0188	5.82	5.82E-09
D_n3_Estabs_TRK	Trucking (terminal or depot)	0.078	0.0117	6.68	2.41E-11
D_n3_Estabs_WDC	WDC	0.113	0.0085	13.3	0
	Number of Retail (NAICS 44-45) Establishments in Buffer				
D_n4445_Estabs_RET	Retail	0.024	0.00417	5.66	1.53E-08
	Number of Trucking (NAICS 484) Establishments in Buffer				
D_n484_Estabs_TRK	Trucking (terminal or depot)	0.089	0.014	6.33	2.39E-10
	Number of WDC (NAICS 493) Establishments in Buffer				
D_n493_Estabs_WDC	WDC	0.062	0.0101	6.12	9.22E-10
	Total Number of Establishments in Buffer				
D_nEstabs_IMX	Intermodal	-0.004	0.0042	-0.955	0.339
D_nEstabs_MAN	Manuf.	-0.028	0.0063	-4.49	7.07E-06
D_nEstabs_REST	Rest	-0.010	0.0020	-5.12	3.04E-07
D_nEstabs_TRK	Trucking (terminal or depot)	-0.017	0.0033	-5.11	3.14E-07
D_nEstabs_WDC	WDC	-0.027	0.0025	-10.5	0
	Likely Fuel / Rest Stop				
E_LikliFuelRest_REST	Rest	1.15	0.09	12.9	0

ACKNOWLEDGMENTS

The submission has been created in part by UChicago Argonne, LLC, Operator of Argonne National Laboratory (“Argonne"). Argonne, a U.S. Department of Energy Office of Science laboratory, is operated under Contract No. DE-AC02-06CH11357. The U.S. Government retains for itself, and others acting on its behalf, a paid-up, nonexclusive, irrevocable worldwide license in said article to reproduce, prepare derivative works, distribute copies to the public, and perform publicly and display publicly, by or on behalf of the Government. Argonne National Laboratory's work was supported by the U.S. Department of Energy, Office of Vehicle Technologies, under contract DE-AC02-06CH11357. This material is based in part upon work supported by the U.S. Department of Energy, Vehicle Technologies Office, under the Systems and Modeling for Accelerated Research in Transportation (SMART) Mobility Laboratory Consortium, an initiative of the Energy Efficient Mobility Systems Program.

AUTHOR CONTRIBUTIONS

The authors confirm contribution to the paper as follows: study conception and design: M. Stinson, O. Sahin; data collection: O. Sahin, M. Stinson; analysis and interpretation of results: Stinson, O. Sahin; draft manuscript preparation: M. Stinson. All authors reviewed the results and approved the final version of the manuscript.

REFERENCES

[1] M. Ruan, J. Lin and K. Kawamura, "Modeling urban commercial vehicle daily tour chaining," *Transportation Research Part E: Logistics and Transportation Review,* vol. 48, no. 6, 2012.

[2] J. Holguín-Veras and G. Patil, "Observed Trip Chain Behavior of Commercial Vehicles," *Observed Trip Chain Behavior of Commercial Vehicles,* vol. 1906, no. 1, 2005.

[3] J. Hunt and S. K. , "Tour-based microsimulation of urban commercial movements," *Transportation Research Part B,* vol. 41, no. 9, p. 981–1013, 2007.

[4] F. Liu, Z. Gao, D. Janssens, B. Jia, G. Wets and Y. Yang, "Identifying business activity-travel patterns based on GPS data," *Transportation Research Part C: Emerging Technologies,* vol. 128, 2021.

[5] H. Borhan, M. Stinson, S. Moura and R. Radulescu, "Connected and Learning Based Optimal Freight Management for Efficiency," in *DOE Vehicle Technologies Office Annual Merit Review (AMR)*, 2022.

[6] B. Sharman and M. Roorda, "Analysis of Freight Global Positioning System Data," *Analysis of Freight Global Positioning System Data, 2246 (1),* 2011.

[7] A. Romano Alho, T. Sakai, M. Chua, K. Jeong, P. Jing and M. Ben-Akiva, "Exploring Algorithms for Revealing Freight Vehicle Tours, Tour-Types, and Tour-Chain-Types from GPS Vehicle Traces and Stop Activity Data," *Journal of Big Data Analytics in Transportation,* vol. 1, no. 2-3, pp. 175-190, 2019.

[8] M. A. Figliozzi, L. Kingdon and A. Tilkitzki, "Analysis of Freight Tours in a Congested Urban Area Using Disaggregated Data: Characteristics and Data Collection Challenges," 2007. [Online]. Available: https://pdxscholar.library.pdx.edu/cgi/viewcontent.cgi?article=1068&context=cengin_fac.

[9] S. You and S. Ritchie, "Tour-Based Truck Demand Modeling with Entropy Maximization Using GPS Data," *Journal of Advanced Transportation,* 2019.

[10] A. Subramanyam, T. Cokyasar, J. Larson and M. Stinson, "Joint Routing of Conventional and Range-Extended Electric Vehicles in a Large Metropolitan Network," *Transportation Research Part C: Emerging Technologies,* 2022.

[11] T. Akter and S. Hernandez, *Truck Activity Pattern Classification Using Anonymous Mobile Sensor Data,* TRB Innovations in Freight Data Workshop, 2017.

[12] X. Yang, Z. Sun, X. Ban and J. Holguin-Veras, "Urban freight delivery stop identification with gps data," *Transportation Research Record: Journal of the Transportation Research Board, 2411 (1),* 2014.

[13] T. Akter and S. Hernandez, "Truck industry classification from anonymous mobile sensor data using machine learning," *International Journal of Transportation Science and Technology,* pp. 522-535, 2022.

[14] M. Ben-Akiva and S. Lerman, Discrete Choice Analysis, Cambridge, Massachusetts: MIT Press, 1985.

[15] M. Ester, H.-P. Kriegel, J. Sander and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," *KDD'96: Proceedings of the Second International Conference on Knowledge Discovery and Data Mining,* vol. 96, p. 226–231, 1996.

[16] S. Hadavi, S. Verlinde, W. Verbeke, C. Macharis and T. Guns, "Monitoring Urban-Freight Transport Based on GPS Trajectories of Heavy-Goods Vehicles," *IEEE Transactions on Intelligent Transportation Systems,* vol. 20, no. 10, pp. 3747-3758, 2019.

[17] A. Thakur, A. R. Pinjari, A. B. Zanjani, J. Short, V. Mysore and S. F. Tabatabaee, "Development of Algorithms to Convert Large Streams of Truck GPS Data into Truck Trips," *Transportation Research Record: Journal of the Transportation Research Board,* vol. 2529, pp. 66-73, 2015.

[18] A. Thakur, A. R. Pinjari, A. B. Zanjani, J. Short, V. Mysore and S. F. Tabatabaee, "Development of Algorithms to Convert Large Streams of Truck GPS Data into Truck Trips," *Transportation Research Record: Journal of the Transportation Research Board, 2529,* 2015.

[19] X. Ma, Y. Wang, E. McCormack and Y. Wang, "Understanding Freight Trip-Chaining Behavior Using a Spatial Data-Mining Approach with GPS Data," *Transportation Research Record: Journal of the Transportation Research Board,* 2016.

[20] K. Gingerich, H. Maoh and W. Anderson, "Classifying the purpose of stopped truck events: An application of entropy to GPS data," *Transportation Research Part C: Emerging Technologies 64,* p. Transportation Research Part C: Emerging Technologies, 2016.

[21] M. Stinson and O. Sahin, "Depot Detection for Freight Tour Data Development," *Accepted for presentation at the TRB Freight Data Innovations Workshop,* 2023.

[22] B. Xu, R. Gupta, B. Hashisho, R. Köhn and S. v. d. Hoef, "Extracting Journeys from Truck GPS Traces," in *IWCTS '22: Proceedings of the 15th ACM SIGSPATIAL International Workshop on Computational Transportation Science*, 2022.

[23] U.S. Census Bureau and the U.S. Department of Commerce, "TIGER/Line Shapefile, 2019, nation, U.S., Primary Roads National Shapefile," 2019. [Online]. Available: https://catalog.data.gov/dataset.

[24] CoStar, "CoStar: Commercial Real Estate Information Company," [Online]. Available: https://www.costar.com. [Accessed 21 July 2020].

[25] Yahoo! Finance, [Online]. Available: https://finance.yahoo.com/. [Accessed 2020].

[26] Bloomberg, [Online]. Available: https://www.bloomberg.com/. [Accessed 2020].

[27] FleetSeek, [Online]. Available: https://fleetseek.com/. [Accessed 2020].

[28] Dun & Bradstreet, [Online]. Available: https://www.dnb.com/. [Accessed 2020].

[29] Google, "Google Maps," [Online]. Available: https://www.google.com/maps. [Accessed 2022].

[30] W. Chang, J. Cheng, J. Allaire, C. Sievert, B. Schloerke, Y. Xie, J. Allen, J. McPherson, A. Dipert and B. Borges, "shiny: Web Application Framework for R. R package," 2022.

[31] M. Bierlaire, "A short introduction to Biogeme. Technical report TRANSP-OR 230620," Transport and Mobility Laboratory, ENAC, EPFL, 2023.

[32] U.S. Department of Transportation, Bureau of Transportation Statistics, "National Transportation Atlas Database," [Online]. Available: https://www.bts.gov/ntad. [Accessed 2023].

[33] J. Hunt and K. Stefan, "Tour-based microsimulation of urban commercial movements," *Transportation Research Part B: Methodological,* vol. 41, no. 9, pp. 981-1013, 2007.