ARTICLE IN PRESS

International Journal of Transportation Science and Technology xxx (xxxx) xxx

Contents lists available at ScienceDirect

International Journal of Transportation Science and Technology

journal homepage: www.elsevier.com/locate/ijtst

Representative truck activity patterns from anonymous mobile sensor data

Taslima Akter [a,⇑], Sarah Hernandez [b]

[a] Infrastructure Analytics Unit, CPCS Transcom Inc., Washington, DC, USA
[b] Department of Civil Engineering, University of Arkansas, Fayetteville, USA

a r t i c l e i n f o

Article history:
Received 29 November 2021
Received in revised form 28 February 2022
Accepted 1 May 2022
Available online xxxx

Keywords:
Freight transportation
Global Positioning System (GPS)
Machine learning
Pattern recognition systems
Trip chaining

a b s t r a c t

With new sources of big data, it is increasingly possible to practically implement advanced freight forecasting models including activity-based and truck touring models. Such models improve upon traditional trip-based approaches by capturing freight behaviors sensitive to transportation policy and infrastructure changes. A persistent challenge with the use of big data in this context is the ability to generalize a set of representative behaviors to serve as the basis for model calibration and validation from anonymized data depicting the complex behaviors of the population. To address this challenge, we present a two-stage methodology to extract unique and representative freight activity patterns from passively collected truck Global Positioning System (GPS) data. The first stage involved a heuristic-based approach to derive a set of stop and trip characteristics from large-streams of GPS pings. The second stage employed data mining and machine learning techniques to discern common freight activity patterns from the set of defined features. The resulting activity pattern profiles, defined as chains of activities and their trajectories over time and space, allow us to maintain the anonymity of the trucks included in the GPS dataset while providing high-resolution travel profiles- a necessary condition for most data sharing agreements between public agencies and private data providers. These activity patterns serve as the critical, and currently missing, data needed to calibrate and validate advanced freight forecasting models. With more advanced forecasting models reflective of observed freight behaviors, we will be able to evaluate a wider spectrum of policy and infrastructure scenarios more accurately.

1. Introduction

Although a number of theoretical Activity Based Models (ABMs) and truck touring models have been developed from as early as 1979 (Adler and Ben-Akiva, 1979), practical implementations have been hindered in part by the unavailability of the data necessary to construct these advanced freight demand forecasting models. In more recent history, growing availability and access to big data from cell phones, Global Positioning Systems (GPS), etc., seemingly closes this data gap. However, we still lack the ability to generalize a set of representative travel patterns from the more complex behaviors of the truck population contained in big data. A representative set of travel patterns is necessary for practical calibration and validation of

Peer review under responsibility of Tongji University and Tongji University Press.
⇑ Corresponding author at: 1028 33rd Street, Suite 320, N.W., Washington, DC, 20007, USA.
E-mail address: takter@cpcstrans.com (T. Akter).

https://doi.org/10.1016/j.ijtst.2022.05.002
2046-0430/

advanced freight travel demand models. Our study develops a methodology to extract unique, representative, and anonymous truck activity patterns from historical truck GPS data, a common source of big data for freight. In this way, we seek to fill a critical research gap concerning the use of big data for advanced freight demand forecasting.

An activity pattern is defined by start/end times, activity duration, travel duration and length, and sequence of those components. Activity patterns have traditionally been derived from travel surveys (Nepal et al., 2005, 2006; Ruan et al., 2012; Allahviranloo et al., 2017) and, less commonly, from mobile sensors (Chung and Shalaby, 2005). Travel surveys have the benefit of linking activity patterns to demographic characteristics but are limited by smaller sample sizes and temporal scopes, e.g., daily or weekly trip diary formats. It can be difficult to extrapolate activity patterns from a one-day travel survey to the population given the complex decision-making processes related to trip chaining. Moreover, travel diaries for freight trucks are almost non-existent. For example, the Vehicle Inventory and Use Survey (VIUS) carried out by the FHWA gathered data from fleet managers on annual trip and vehicle characteristics but did not at all resemble a typical trip diary that was needed to recreate travel patterns (FHWA, 2001).

Counter to travel surveys, passively collected mobile sensor data for freight captures a much larger proportion of the truck population and provides continuous spatial and temporal coverage. This data is increasingly available due to the prevalence of on-board or cellphone-based GPS units and, recently mandated, Electronic Logging Devices (ELD). Since mobile sensor data typically represents a large but sampled portion of the population, it has been commonly used as a source of probe vehi-cle data to measure speeds and travel times. Considering this data depicts high resolution vehicle movements, sometimes on the order of minute-to-minute position updates, and is potentially available for all trucks, there is a significant power in leveraging it to gain insights into freight activity patterns.

Although the private sector collects large sample of robust data, the publicly shared data are void of industry, commodity, fleet, and driver information due to business privacy concern. Thus, a persistent challenge with the use of freight big data in this context is the ability to generalize a set of representative behaviors to serve as the basis for model calibration and validation. Also, there is a critical research gap to identify commodity-specific activity patterns from the anonymized data depicting the complex behaviors of freight population while maintaining data privacy. To address this challenge, we present a two-stage methodology to extract unique and representative freight activity patterns from passively collected truck Global Positioning System (GPS). Our approach identifies commodity clusters of freight trucks based on their daily activity patterns in a way that maintain business privacy standards.

The first stage involved a heuristic-based approach to derive a set of stop and trip characteristics from large-streams of GPS pings. The second stage employed unsupervised machine learning techniques, namely K-means clustering, to discern common freight activity patterns from the set of defined features. The premise of this study follows from the work of Allahviranloo et al. (2017) for passenger activity travel pattern generation. Allahviranloo et al. (2017) demonstrated, using survey data, that a limited set of representative daily activity patterns can be extracted from those of the larger population and used for ABM calibration and validation. Our work not only extends this approach to freight activity pattern recognition but leverages anonymous mobile sensor data in place of traditional travel surveys.

2. Background

Trucking is and will continue to be the dominate mode of transport for freight in the US with trucks accounting for 64% and 69% of the market by both weight and value, respectively (FHWA, 2018). The Freight Analysis Framework (FAF), the Federal Highway Administration's (FHWA) nationwide freight forecasting model estimates that the weight of freight shipments moved by truck will grow 45% between 2012 and 2045 (FHWA, 2018). Ensuring efficient freight movement through the provision of adequate infrastructure and effective transportation policy is critical for the economy and the environment. To construct, maintain, and operate a transportation system that supports the efficient movement of freight, it is necessary for public transportation agencies accurately model and predict freight travel demands.

A variety of travel demand models, i.e., traditional trip-based, activity-based, and truck touring models, are used to predict freight flows and, in turn, direct effective freight-oriented infrastructure and policy programs. However, the choice of an appropriate model depends on data availability, time and resource allotments, and the need to assess certain infrastructure and/or policy scenarios. Advanced freight forecasting models are increasingly used to predict travel demands as they consider robust behavioral characteristics, operational decisions, and interactions. Advance models, compared to their traditional trip-based predecessors, allow agencies to evaluate a wider variety of infrastructure and policy decisions by incorporating behavioral models. Activity Based Models (ABMs), for example, forecast travel demand by depicting trip chains of individual agents participating in a set of activities. For freight, activities include initiating/receiving shipments and transporting goods from origin to destination by various modes. Agents may be shippers, receivers, or drivers. The premise of such models, unlike trip-based models, is that travel is derived from the demand to pursue activities. Thus, models that consider trip linkages have the potential to more accurately forecast travel demands by focusing on activity patterns rather than individual trips.

With new sources of big data providing insights into freight travel patterns, it is becoming increasingly possible to practically implement advanced freight forecasting models including activity-based and truck touring models. Key to successfully leveraging big data for advanced travel demand modeling is the ability to (1) derive operational characteristics, (2) extract common activity patterns, and (3) link activity patterns to the population.

2.1. Deriving operational characteristics

In order to distill common activity patterns from big data sources like GPS, it is first necessary to extract operational characteristics that define activity patterns. Examples of operational characteristics include trip length, number of trips, speed, travel time, destination, stop location, and stop duration (Zanjani et al., 2015; Liao, 2009; Gingerich et al., 2016; Aziz et al., 2016, Laranjeiro et al., 2019). Heuristic approaches for identifying stops (''Stop Identification'') and trips (''Map Matching'') have been developed to derive operational characteristics from large-streams of GPS data (Giovannini, 2011; Thakur et al., 2015; Quddus and Washington, 2015; Camargo et al., 2017; Yang et al., 2014). ''Stop-identification'' refers to finding clusters of pings that relate to a single stop. Available algorithms (Thakur et al., 2015; Camargo et al., 2017) used geographic bound-ing boxes and rule-based approaches to define stop clusters. ''Map-matching'' refers to the process of identifying the network links that correspond to each GPS ping (a latitude, longitude, timestamp tuple). Yang et al. (2014) developed a support vector machine (SVM) to identify delivery stops with GPS data. The features of the SVM model were the stop duration, the distance from a stop to the center of the city, and the distance to a stop's closest major bottleneck.

Oka et al. (2019) developed a route choice model using the large-scale urban freight survey and found that travel patterns changed significantly depending on the type of trucks. Giovannini (2011) developed an algorithm to re-construct routes from low-sample rate GPS data, e.g., around one mile between pings, using a Bayesian approach (Giovannini, 2011). Quddus and Washington (2015) developed a new weight-based shortest path and vehicle trajectory aided map-matching algorithm to determine the network link corresponding to each GPS ping based on proximity, among other factors, for a sparse road network. Further extensions of map-matching, such as that by Camargo et al. (2017), ensured that the sequence of identified network links constituted a complete path. The ''stop identification'' and ''map-matching'' algorithms developed by Camargo et al. (2017) were used in this paper as they were shown to produce accurate stop locations and routes for GPS data. We applied several modifications to their algorithms to ensure accuracy for denser road networks and less urbanized areas (see Section 3.1).

2.2. Extracting representative activity patterns

Due to the ability to handle complex patterns and noise found in large datasets, machine-learning techniques have been used to extract representative activity patterns from surveys (Allahviranloo et al., 2017; Jiang et al., 2012; Allahviranloo and Recker, 2013; Li and Lee, 2017) and mobile sources (Shoval and Isaacson, 2007; Yang et al., 2010; Liu et al., 2014). Jiang et al. (2012) applied Principle Component Analysis (PCA) and K-means clustering to extract representative groups among weekday and weekend activity patterns from travel surveys. They found eight and seven representative groups for weekdays and weekends, respectively. Allahviranloo and Recker (2013) used Support Vector Machines (SVM) to classify the daily activity patterns of travelers based on trip diary data. Allahviranloo et al. (2017) generated activity patterns from survey data using a combination of affinity propagation and K-means clustering. They defined 12 activity patterns, where the pattern corresponding to long duration work activity was the most prevalent. Also working with survey data, Li and Lee (2017) developed a Probabilistic Context Free Grammar (PCGG) model to analyze and generate daily activity patterns. They found 15 common activity patterns which explained 70% of the behaviors represented by their data sample.

Ma et al. (2016) used a series of data-mining algorithms to extract an individual truck's trip-chaining information from multiday GPS data. You et al. (2016) developed a modeling framework for freight that considered both spatial–temporal constraints. They developed the model based on an adaptation of an activity-based passenger model called the Household Activ-ity Pattern Problem (HAPP). Hunt and Stefan (2007) developed a tour-based microsimulation model using a set of interviews about own-account commercial vehicle movements conducted at just over 3100 business enterprises in the Calgary Region. Joubert and Meintjes (2015) argued that the use of GPS data and the associated expert judgements can be applied with confidence in freight transport models.

Shoval and Isaacson (2007) used a variety of tracking technologies, i.e., GPS tracking, Cellular Triangulation tracking, assisted GPS tracking, and land-based time difference of arrival (TDOA) tracking, to collect and analyze time–space activity patterns of tourists. They found that GPS devices collected more accurate data than other tracking methods. Like the studies by Allahviranloo and Recker (2013) and Allahviranloo et al. (2017), Yang et al. (2010) applied SVM methods to determine the individual's travel behavior but used GPS data instead of travel surveys. Features used to train their SVM included activity start time, end time, distance, etc. derived from the GPS data (Yang et al., 2010). They were able to distinguish around eight unique activity patterns. Similarly, Liu et al. (2014) used mobile phone data to identify activity types based on travel behaviour information, i.e., the timing and frequency of visits to different locations. Liu et al. (2015) developed a model based on profile Hidden Markov Models (pHMMs) to quantify the occurrence probabilities of all the daily activities as well as their sequential order also using mobile sensor data. They found three main patterns dependent on the location of the longest activity duration, i.e., home, work, and non-work clusters, where the non-work cluster had seven sub-clusters. Considering the availability of truck GPS data, there is significant potential in extending the abovementioned techniques to distill activity patterns for freight.

2.3. Linking representative activity patterns to the population

To expand representative activity patterns extracted from surveys or samples of mobile sensor data to the population-at-large, it is necessary to link patterns to freight demographic characteristics like industry served and commodity carried. However, commercially available mobile sensor data is typically devoid of demographic data, e.g., anonymized, to protect the privacy and satisfy data sharing agreements between public agencies and private data providers. Jing (2018) attempted to overcome this limitation by concurrently collecting travel diary and GPS data for freight trucks through a tablet-based application. Like traditional travel surveys, this approach was restricted by its smaller sample size (i.e., the survey included only 119 truck drivers in Singapore), bringing into question the ability to extrapolate derived activity patterns to a much larger truck population (Jing, 2018).

Sharman and Roorda (2013) used truck GPS data to calculate the time between arrivals at a destination of two successive vehicles operated by the same carrier and found a wide variation in shipping behavior of commercial establishments. They argued that many firms do not follow consistent shipping schedules. Alho et al. (2019) compared different stop-to-tour assignment, tour-type, and tour-chain identification algorithms to extract their implications. They demonstrate that the predictions of tours, tour types, and tour-chain-types are highly dependent on the used assumptions and GPS data processing. They also analyzed the daily tour pattern and demonstrated that the operational characteristics varies from vehicle types and industries. You and Ritchie (2018) analyzed truck GPS data to interpret tour behavior of clean drayage trucks, and to prepare sufficient tour data for clean truck modeling. Sánchez-Díaz et al. (2015) developed a freight demand model that estimated tour flows from secondary data sources e.g., traffic counts and bypass the need for expensive surveys.

Without survey data to provide necessary demographics like trip purpose, commodity carried, or truck type, algorithmic approaches to derive such information from GPS data have been attempted. Kuppam et al. (2014) combined GPS and land use data to derive trip purposes, i.e., goods pickup or delivery, service, return home. They showed that land use at the trip origin was a significant predictor of trip purpose and was able to correlate industry type with trip characteristics like fre-quency and number of stops. For example, ‘‘construction trucks” made fewer stops than ‘‘government-related trucks”. Unlike the study by Kuppam et al. (2014) which was able to correlate freight demographics from activity or trip characteristics, Ma et al. (2011) focused on distinguishing vehicle characteristics from mobile sensor data, which can also be useful for inferring freight demographics. They used GPS data to classify truck trips into access, local, and loop trips based on trip travel distance from the origin to the destination relative to straight-line distances. Similar to these approaches, the methodology described in this paper connects activity patterns to freight demographics, specifically industry served, by examining land uses at each stop location.

3. Methodology

Following a brief discussion of the data requirements, the two major components of the methodology are discussed in this section: (1) derivation of operational characteristics from truck GPS data, and (2) selection, estimation, and validation of unsupervised machine-learning models to discern unique truck activity patterns from operational characteristics.

3.1. Data requirements

The methodology described in this paper is suited for large streams of mobile sensor data that contain a unique, but anonymous, vehicle identification number (ID), timestamp, latitude and longitude, point-speed, and heading direction (e.g., azimuth). GPS data used in this paper is in a raw form that required pre-processing to remove noise and other inconsistencies in the data. It is assumed that adequate quality checks will produce ‘complete’ truck records, defined as those that represent an over the road truck movement with reasonable start and end positions, speeds, and accelerations. We developed an algorithmic approach to identify a complete truck tour after checking data consistency and relevancy. The algorithm removes any GPS pings that have acceleration/ deceleration rates above 2.24 ft/s2, corresponding to 85th percentile average acceleration rate of heavy trucks (Pline, 1999). It also flags any truck records that have space-mean-speed (SMS) more than 81 miles per hour (mph) for at least 2 min. Additionally, the algorithm calculates the spatial and temporal extension of truck records to check the validity of the data. Any truck records that have a smaller geographic and temporal coverage than the thresholds are removed. The spatial and temporal thresholds (i.e., 1.2 miles and 20 min) were determined by observing manually the spatial extent of freight distribution centers and operational stops (e.g., pickup/ delivery, rest stops, etc.).

Once cleaned of inconsistencies, GPS data represented as a series of pings should be converted to a series of stops and trips. Heuristic approaches developed by Camargo et al. (2017) to identify stop clusters and routes from truck GPS data were adapted for this work due to differences in proposed application contexts, i.e., metropolitan area vs statewide region. To define stop location, we used the first identified ping in the stop cluster (e.g., a group of consecutive pings with minimal speed) rather than using the cluster centroid as the stop location. This ensured that stop locations aligned with physical business locations. First, the ‘‘stop identification” algorithm calculates the space-mean speed (SMS) from consecutive GPS pings. It continues calculating the SMS between the next pair of consecutive pings until the SMS is more than 3 mph. Thus, the algorithm collects an array of the pings that pass the speed threshold (i.e., less than or equal 3 mph). Next, the algorithm calculates the total ‘‘stop time” and ‘‘stop coverage” for all consecutive pings from the array. To avoid traffic signal related stops, we added a spatial and a temporal threshold to the algorithm. Thus, if a group of stopped pings covers at most 0.2 miles for at least 5 min, then the group is considered as a ''stop cluster'' and the first ping of that ''stop cluster'' is marked as the location of the ''stop''. Later, these stops are categorized as pick-up/delivery stops, rest or fuel stops, or unintended stops due to congestion.

After categorization of stops, we developed the ''path identification'' algorithm to identify the set of links that comprised the complete path between consecutive pings and extract the trip characteristics. In regard to trip characteristics, modifications to the ''map-matching'' algorithm by Camargo et al. (2017) accounted for a dense statewide road network. Use of the All Roads Network of Linear Referenced Data (ARNOLD) (FHWA, 2014) network file in this work, ensures the transferability of results from state-to-state. Because this network was denser than that used by Camargo et al. (2017), the link buffer distance was altered based on road functional class to improve accuracy in matching GPS pings to network links. Additionally, the modified algorithm defined link cost using estimated free-flow travel time instead of link length. Since ARNOLD does not include speed limits, speed limits were assumed based on road functional class. Further details on modifications to the stop identification and map-matching algorithms can be found in Akter et al. (2018).

3.2. Operational characteristics as input feature vector

Five operational characteristics were extracted from the GPS data, three relating to stops, i.e., stop time of day, number of stops, and stop duration and two relating to trips, i.e., trip length and trip duration. The extracted stop duration was divided into three groups based on the Hours of Service (HOS) regulations of the Federal Motor Carrier Safety Administration (FMCSA). According to the ''30-Minute Driving Break'' rule of HOS, all freight drivers are required to take a 30-minute break after an 8 h cumulative drive (FMCSA, 2020). The ''14-Hour Limit'' rule regulates a consecutive 10 h break for freight drivers after having been on duty for 14 h. Additionally, drivers may split their required 10-hour off-duty period between at least 2 and 8 h while using sleeper berths (FMCSA, 2020). Hence, we considered '30 min to 8 h' as the critical threshold for extract-ing activity patterns.

To derive daily activity patterns, we segmented multi-day travel patterns by day (i.e., from midnight to midnight). For instance, if a unique truck traveled for three days, that truck would be segmented into three independent daily truck records. We adopted this approach to consider situations where a unique truck transported different goods on different days and thus showed different activity patterns. The daily pattern of each truck was represented by an 11-element feature vector based on operational characteristics (Table 1).

These features relate to behavioral characteristics assumed to distinguish representative activity patterns. For instance, stops of 'less than 30 min' duration captured short-breaks, e.g., food break, restroom, refueling, short-term deliveries etc. while stops of '30 min to 8 h' duration captured long-term pickup/delivery stops but not long rest periods (Jing, 2018). Trip length and trip duration were used to identify the types of truck trips. Trip lengths 'less than 30 miles' and/or trip duration 'less than 1 hour' were assumed to represent short-haul truck movements while trip lengths 'more than 100 miles' and/or trip duration 'more than 4 h' represented long-haul truck movements.

3.3. Unsupervised machine learning to derive representative activity patterns

A K-means clustering model was applied to identify the representative daily activity patterns of trucks. The assumption was that K-means clustering could distill the daily activity patterns of the truck population to a relatively small set of representative patterns, as well as to identify the optimal number and compositions of such patterns should they exist (Allahvira et al., 2017).

Unsupervised learning methods find multi-dimensional groups in data represented by multi-dimensional input vectors (Alpaydın, 2014). Among unsupervised learning models (i.e., Hierarchical, DBSCAN, Gaussian Mixture Model, etc.), K-means cluster models are appropriate when input variables are numerical, as is the case for the feature vector representing operational patterns (Bishop, 2016). K-means clustering algorithms partition the data into K number of clusters in a

Table 1
Features defined by operational characteristics by group and type.

Feature Group	Features	Variable Type
Stop Duration	1. Number of stops less than 30 min 30 min to 8 h More than 8 h	Discrete
Trip Length	2. Number of trips less than 30 miles 30 miles to 100 miles More than 100 miles	Discrete
Trip Duration	3. Number of trips less than 1 hour 1 hour to 4 h More than 4 h	Discrete
Time of Day	4. Proportion of daytime stops (6AM – 6PM)Proportion of nighttime stops (12AM – 6AM & 6PM – 12AM)	Continuous

multidimensional space such that the sum of the squares of the distances of each data point to its closest cluster centroid μ_k is a minimum (Bishop, 2016) (Eq. (1)). A two-step iterative procedure is used to find optimal cluster assignments. Iterations correspond to successive optimizations with respect to the binary indicator variables for cluster membership (r_{nk}) and the cluster centroid ''location' (μ_k). The first step assumed a random value for μ_k for K number of clusters and minimizes J with respect to r_{nk} (Eq. (2)). In the second step, J is minimized with respect to μ_k, keeping r_{nk} fixed (Eq. (3) and (4)). The first stage of updating r_{nk} and the second stage of updating μ_k correspond respectively to the E (expectation) and M (maximization) steps of the EM algorithm. This two-stage optimization is repeated until convergence (Bishop, 2016).

$$J = \sum_{n=1}^{N} \sum_{k=1}^{K} r_{nk} \|x_n - \mu_k\|^2 \tag{1}$$

$$r_{nk} = \begin{cases} 1 & \text{if } k = \arg\min_j \|x_n - \mu_j\|^2 \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

$$2 \sum_{n=1}^{N} r_{nk} (x_n - \mu_k) = 0 \tag{3}$$

$$\mu_k = \frac{\sum_n r_{nk} x_n}{\sum_n r_{nk}} \tag{4}$$

Where,

$\{x_1, \ldots, x_n\}$ = N observations of a random D-dimensional Euclidean variable x.
μ_k = Centers of the clusters, where $k = 1, \ldots, K$.
r_{nk} = Binary indicator variables, {0, 1} describing which of the K clusters the data point x_n is assigned to, where $k = 1, \ldots, K$.

A challenge in applying K-means clustering is the need to define the number of clusters when there is no a priori knowledge of appropriate value. Several approaches are suggested in the literature to select K including i) by the rule of thumb, ii) 'elbow' method, iii) information criterion approach, iv) an information theoretic approach, v) choosing K using the silhouette and vi) cross-validation (Kodinariya and Makwana, 2013). Of these methods, the 'elbow' method is the most commonly used and, in this study, produced a logical K value (Ng, 2012). The ''elbow" method considers the number of clusters K as a func-tion of the total within-cluster sum of squares (WSS). A reasonable number of clusters K differences when there is minimal change in the total WSS after adding another cluster.

4. Results

Four, two-week periods of anonymous truck GPS data representing each quarter of the year (i.e., February, May, August/September, and November) gathered from the American Transportation Research Institute (ATRI) were used to assess the proposed method. The data from the August/September sample was used for algorithm calibration, i.e., setting the stop identification and map-matching parameters and determining an appropriate number of clusters, while the remaining datasets were used for assessing temporal transferability. In total, there were approximately 338,304,135 pings within the eight-week sample period. The sample represented 358,092 unique trucks in Arkansas and was shown to be a representative sam-ple of the total truck population of the state (Corro et al., 2019).

The K-means clustering model was applied to approximately 300,000 daily truck movement records and produced six distinct clusters (K = 6) from the 11-element input feature vector (Table 2). The number of clusters (K) was varied from one to 15 clusters and the ''elbow" method was applied to determine a reasonable number of clusters (Fig. 1). Since the WSS plateaued beyond six clusters, minimal differences in cluster characteristics were observed when more clusters were added. Alternatively, total WSS increased when the number of clusters decreased below six clusters.

The following definitions were adopted to facilitate interpretation of activity patterns represented by each cluster:

Short break: Stop duration less than 30 min
Pickup/delivery: Stop duration 30 min to 8 h
Long rest break: Stop duration more than 8 h
Short-trip length: Trip length less than 30 miles
Medium-trip length: Trip length 30 miles to 100 miles
Long-trip length: Trip length more than 100 miles
Short-trip duration: Trip duration less than 1 hour
Medium-trip duration: Trip duration 1 hour to 4 h
Long-trip duration: Trip duration more than 4 h
Daytime hours: 6 AM – 6 PM
Nighttime hours: 12 AM – 6 AM and 6 PM – 12 AM

Table 2
Centroids of K-means Clusters.

	Features	Activity Pattern 1	Activity Pattern 2	Activity Pattern 3	Activity Pattern 4	Activity Pattern 5	Activity Pattern 6
Stop duration	1. Less than 30 min	2 (7.7)	1 (1.5)	0 (0.6)	1 (2.3)	1 (0.7)	1 (1.4)
	2. 30 min to 8 h	3 (5.9)	1 (2.1)	0 (0.5)	1 (1.3)	1 (0.5)	1 (0.8)
	3. More than 8 h	1 (0.3)	1 (0.4)	1 (0.0)	1 (0.0)	0 (0.0)	0 (0.0)
Trip length	4. Less than 30 miles	3 (14.9)	1 (3.0)	0 (1.0)	1 (4.3)	0 (0.8)	1 (1.8)
	5. 30 to 100 miles	2 (2.9)	1 (1.1)	0 (0.5)	1 (0.8)	0 (0.3)	0 (0.5)
	6. More than 100 miles	1 (1.1)	1 (0.8)	1 (0.4)	1 (0.5)	1 (0.3)	1 (0.4)
Trip duration	7. Less than 1 hour	4 (15.6)	1 (3.3)	1 (1.2)	1 (4.5)	0 (0.9)	1 (2.0)
	8. 1 to 4 h	2 (2.2)	1 (1.3)	1 (0.6)	1 (0.9)	1 (0.5)	1 (0.6)
	9. More than 4 h	0 (0.3)	0 (0.4)	0 (0.3)	0 (0.3)	0 (0.3)	0 (0.2)
TOD	10. Day proportion	0.72 (0.01)	0.45 (0.01)	0.00 (0.00)	1.00 (0.00)	0.00 (0.00)	1.00 (0.00)
	11. Night proportion	0.28 (0.01)	0.55 (0.01)	1.00 (0.00)	0.00 (0.00)	1.00 (0.00)	0.00 (0.00)
Percentage of trucks within each activity pattern cluster		9%	11%	14%	20%	14%	32%

Note: The standard deviation of the feature within the samples in the cluster is shown in parenthesis.

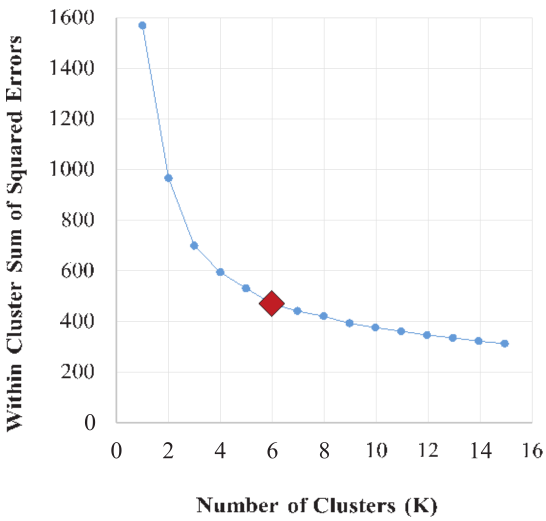


Fig. 1. Number of clusters based on ‘‘elbow method”.

The highest percentage of sampled trucks (about 32%) were clustered into Activity Pattern 6 that had one or two daily stops, specifically during daytime hours. Those stops, either a short break or a pickup/delivery, were followed by both short- and long-trip lengths. The second highest percentage (about 20%) of sampled trucks were grouped into Activity Pattern 4. Those trucks had one to five daily stops (i.e., short break, pickup/delivery, and long rest break) followed by short-, medium-, and long-trip lengths.

All stops in Activity Pattern 4 occurred during daytime hours. Also, we observed that around 14% of sampled trucks were clustered into both Activity Pattern 3 and Activity Pattern 5, independently. Trucks of Activity Pattern 3 had long rest breaks during nighttime hours followed by long-trip lengths. Alternatively, trucks of Activity Pattern 5 had long-trip lengths with no long rest break. Around 11% of sampled trucks in Activity Pattern 2 had one to four daily stops. Those stops were followed by short- and medium-trip durations. Around 55% of stops in Activity Pattern 2 occurred during nighttime hours. Further, we found that about 9% of trucks were clustered into Activity Pattern 1 and had a high number of daily stops (on average 6 stops in a day). Around 33% of those stops were short breaks and 17% were long rest breaks. Moreover, most of the stops (about 72%) occurred during daytime hours for Activity Pattern 1.

5. Discussion

The six representative activity patterns found via K-means clustering using an 11-feature vector depicting operational characteristics can be described according to their spatio-temporal characteristics (Table 3). Time-space diagrams depicting changes in location along the horizontal axis (blue lines), duration of activities and travel along the vertical axis (dashed red lines), and portions of the trip that are unknown (grey wavy lines) (Fig. 2) show distinct patterns. When a trip returned to the same starting location after completing its daily activity, we categorized that trips as a ''home-base" trip. For example, 'Short-Haul Home-Base with Multiple Stops' (e.g., Activity Pattern 1) showed a pattern in which trucks made multiple numbers of stops and returned to their home-base at the end of the day. Trucks labeled 'Medium-Haul Home-Base with One/Multiple Stops' (e.g., Activity Pattern 5) started driving midday after a long rest-break (about 11 h) followed by a series of short breaks and medium-trip durations (Fig. 2a). At the end of the day, those trucks also returned to their assumed home base. The last example, labeled 'Long-Haul with One Stop' (e.g., Activity Pattern 6) showed a pattern in which trucks drove through the night and took a short break at 6 AM before resuming their drive across the state (Fig. 2a). Unlike short and medium-haul movements, these trucks did not return to a home-base by the end of the day. The grey lines represented unknown portions of the trip. This occurred due to the data sample restriction to truck movements within the state boundary. The remaining activity patterns differed in their number and duration of stops, travel distances, and returns to a home base (Fig. 2b). As mentioned earlier, Activity Pattern 2 was similar to 'Short-Haul Home-Base with Multiple Stops' while Activity Patterns 3 and 4 were similar to 'Medium-Haul Home-Base with One/Multiple Stops'.

Key to the uniqueness of the six activity patterns in the study was the definition of the feature vector representing the operational characteristics of the trucks. Stop and trip characteristics were two basic operational characteristics that likely varied by commodity carried and industry of the truck. For example, since early morning is the best time to feed hens, trucks carrying chicken feed make multiple short breaks in the morning (before sunrise) followed by short-trip lengths (Waldroup and Hellwig, 2000). Some industries, like mining, operate 24 h a day and result in a high number of stops and trips through-out the day. By including features that relate to the time of day, stop duration, trip length, and trip duration, we are able to capture these differences in operation that lead to different activity patterns.

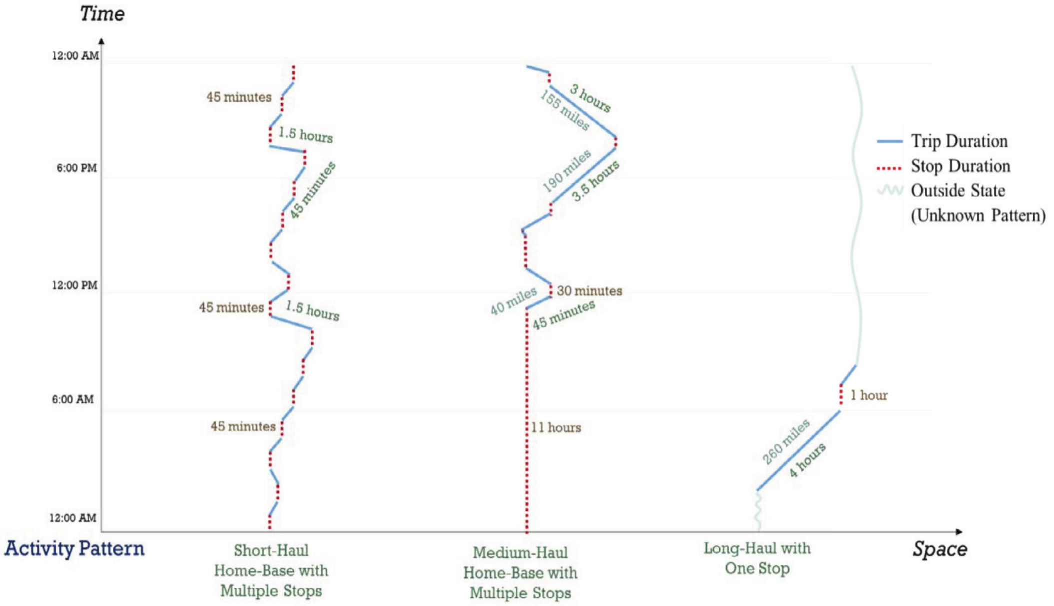
A drawback of K-means clustering is the a priori need to define the number of clusters. To demonstrate the sensitivity of activity patterns to the selected number of clusters, we examined the activity patterns under assumptions of five (K = 5) and seven (K = 7) clusters and noted the trends in cluster centroid definitions as we increased the number of clusters beyond seven. With five clusters, Activity Pattern 5 merged with Activity Pattern 3. Thus, we were unable to see subtle differences in medium-haul trips. Specifically, Activity Pattern 3 had one long-trip duration stop while Activity Pattern 5 had one short-trip duration followed by a pickup/delivery. Increasing the number of clusters from five to six allowed us to distinguish Activity Pattern 5 and Activity Pattern 3. Increasing from six to seven clusters, on the other hand, divided Activity Pattern 1 into two clusters. However, the newly created pattern had no meaningful characteristics that would distinguish it as a unique pattern, only a difference in the number of daily stops without changes in the trip length/duration or sequencing among stops. Thus, six clusters were assumed to capture unique and representative activity patterns from the sample.

Variation in the representative activity patterns arose not only due to the selection of the number of clusters but was also found within the samples that comprised each cluster. Activity Pattern 1, which represented the lowest percent (about 9%) of daily truck samples, had the highest within-cluster variance for each feature. Other activity patterns had relatively smaller within-cluster variation for each feature. Features with the highest within-cluster variation across all clusters included trips less than 30 miles (feature #4) and trip duration less than 1 hour (feature #7) while the lowest variation was found with stop duration more than 8 h (feature #3), trips more than 100 miles (feature #6), and trips longer than 4 h (feature #9). The higher number of short-trips in a day (versus one long-trip) was likely responsible for this variation. High variation among features in Activity Pattern 1 explained why increasing the number of clusters leads to further separation of that pattern. However, the current K-means clustering feature set considers the time series of trips and stops visited by a truck but not the sequence. The representative activity pattern is considered from all trips and stops made by a truck but does not consider the order of trips and stops. This chosen representation within the model can lead to some of the misclassifications. Future work will consider the sequence and frequency of trips and stops to improve on the misclassification.

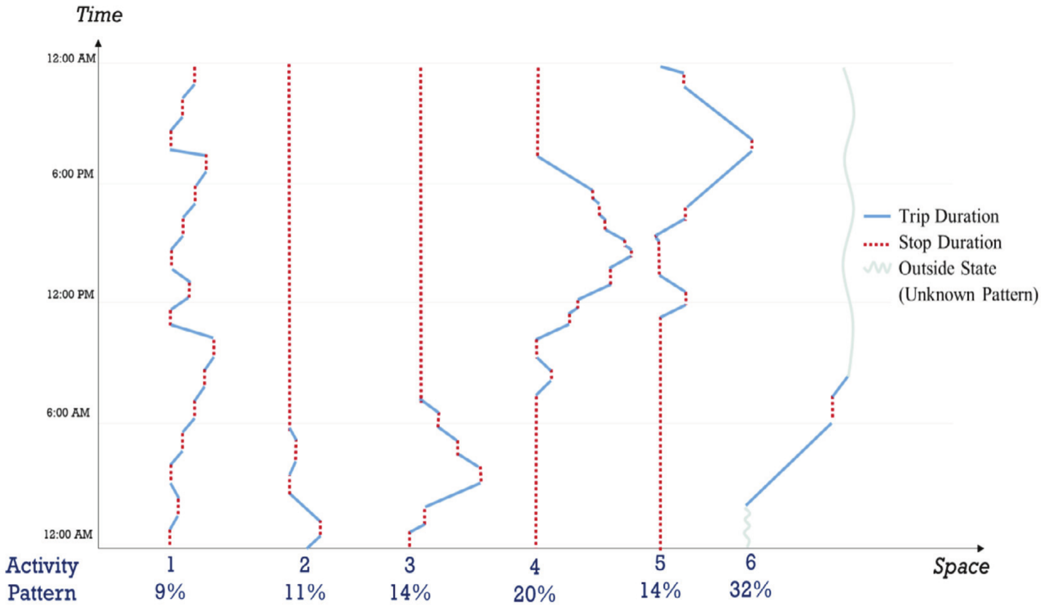
To tie activity patterns distilled from the GPS data sample to those of the larger population for which demographics are known, it was necessary to link each pattern to freight demographics such as commodity or industry type. To create this

Table 3
Categorization of activity patterns.

Activity Pattern	Category Name	Category Description
Activity Pattern 1 Activity Pattern 2	Short-Haul Home-Base with Multiple Stops	Trucks have multiple stops followed by multiple short trips and return to home-base within a day
Activity Pattern 3 Activity Pattern 4 Activity Pattern 5	Medium-Haul Home-Base with One/ Multiple Stops	Trucks have one/multiple stops followed by one/multiple medium trips and return to home-base within a day
Activity Pattern 6	Long-Haul with One Stop	Trucks have one (or two) stop followed by one long trip and not return to home-base within a day



(a) Examples of activity pattern types



(b) Activity pattern examples for six clusters

Fig. 2. Daily activity patterns of freight trucks.

linkage, 2064 daily activity patterns were mapped using Google Earth, and the business types of each stop location were examined to determine the industry served by the truck. We inspected truck GPS stop locations using aerial imagery of land use and business locations (e.g., Google Satellite images) and created a set of industry labeled data. Since the satellite imagery includes building locations, orientations, access roads, loading docks, and other details, we can get the insights of which freight industry is served by that truck. We followed a sequential manual inspection that involved observing the identified stop location, stop time, and stop duration of a truck against satellite imagery (Google Earth), examining type of businesses at the stop using web searches, predicting the industry served by that truck, and comparing the predicted industry served

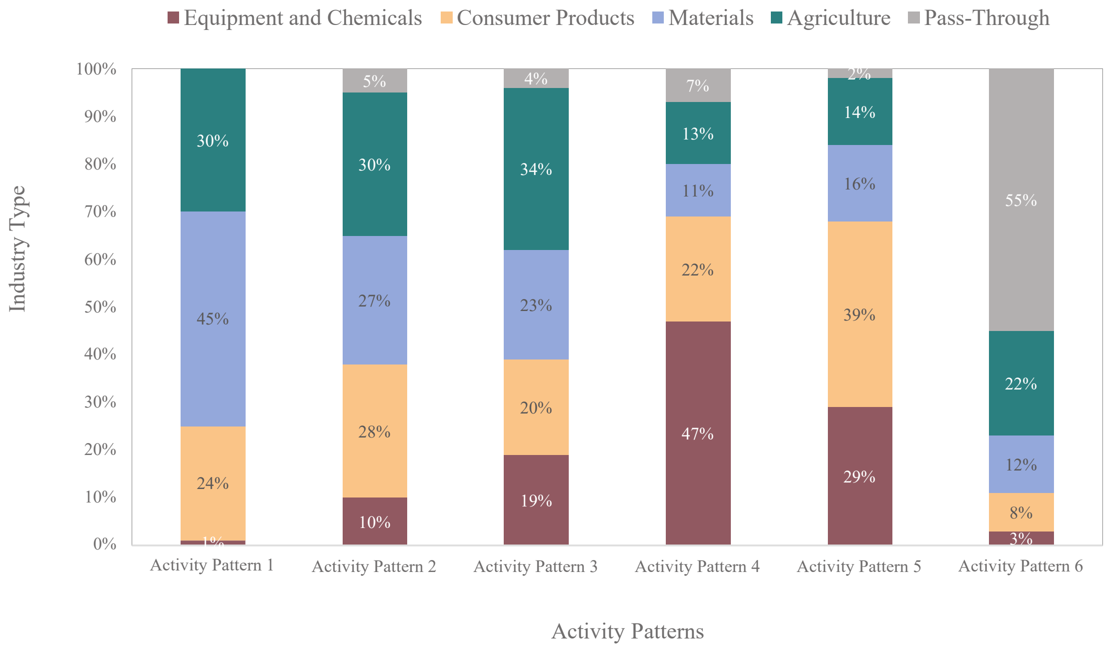


Fig. 3. Industry types contained in each activity pattern cluster.

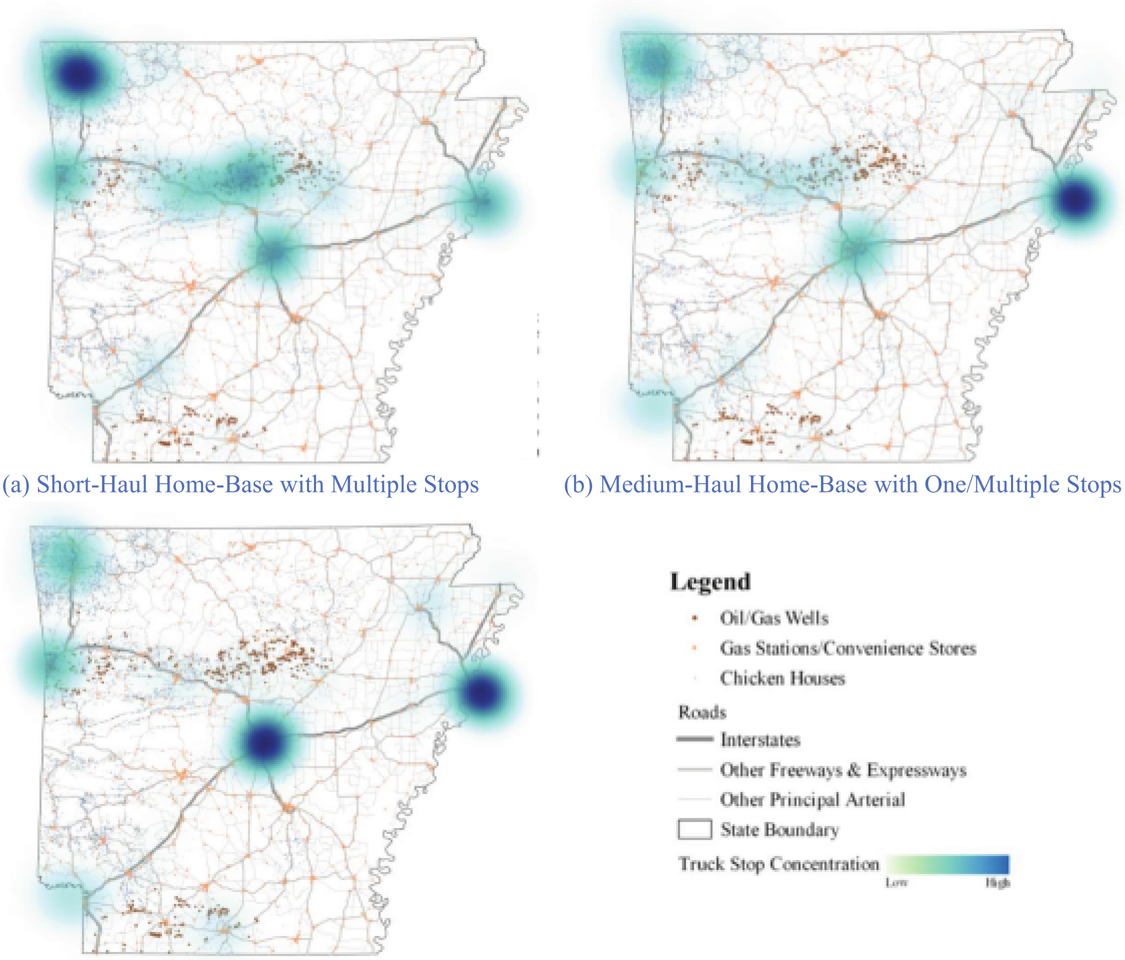
against the clustering model's activity patterns. Thus, we inferred industry types for 2,064 daily truck records. Industry types were aggregated into five groups defined as follows:

(1) Agriculture including agriculture and livestock
(2) Materials including mining, coal, oil/gas, and non-metallic minerals
(3) Consumer products including food, lumber, and other manufactured products
(4) Equipment and chemicals including paper, chemicals, concrete, and metals
(5) Pass-through which included stops at rest areas and gas stations

Each activity pattern cluster consisted of trucks serving multiple industries, however, there was a dominant industry group for several of the activity patterns (Fig. 3). Of all trucks included in Activity Pattern 1, 45% served the materials industry and 30% served the agriculture industry (Fig. 3). We assumed this was in line with operations of trucks traveling to and from oil and gas wells to support fracking activity, e.g., many short duration stops and trips with a return to a home base at the end of the day. Further supporting this assumption was the location of stops for Activity Pattern 1 (i.e., Short-Haul Home-Base with Multiple Stops) which align with known oil and gas wells (Fig. 4a). Those same locations also had businesses related to poul-try which tend to generate short-haul truck trips between feed mills, chicken houses, and processing facilities (e.g., chicken houses). Activity Patterns 2 and 3 shared similar distributions among industry types with agriculture representing approxi-mately 30 and 34%, followed by materials representing approximately 27 and 23%, respectively (Fig. 3).

Activity Patterns 3 was distinguished by several medium-trip lengths and short breaks with a return to a home base. We assumed the activity patterns related to agriculture in this case were capturing grain production and processing where movements were within the state (i.e., Medium-Haul Home-Base with One/Multiple Stops) to and from farms and centralized grain elevators. This was also seen in the relatively heavier volumes of Activity Pattern 3 trucks in the northeast and north-west regions of the state where farms are located (Fig. 4b). For materials, we assumed the medium-haul, home based activ-ities captured movements of petroleum between fueling stations. Further, about 55% of trucks following Activity Patterns 6 represented pass-through movements (Fig. 3). The heatmaps of Activity Pattern 6 (i.e., Long-Haul with One Stop) also showed that these trucks had a high concentration of stops in the center region of the state (Fig. 4c). We considered this pattern as pass-through truck movements that took short-breaks followed by long-trip lengths. The approach of linking activity pattern to industry type is transferable to any geographic extent, although industry types may differ based on the area.

However, due to the anonymity of GPS data, it was not possible to directly ''observe" the demographic characteristics of the trucks within each representative pattern. The unsupervised model (i.e., K-means clustering) of this paper identified unique truck activity patterns that can support transportation agencies to develop advanced freight forecasting models. Future work will consider applying supervised machine learning techniques to predict industry-served or commodity-carried of freight trucks from the extracted operational characteristics.



(a) Short-Haul Home-Base with Multiple Stops

(b) Medium-Haul Home-Base with One/Multiple Stops

(c) Long-Haul with One Stop

Fig. 4. Stop location concentration by activity pattern.

6. Conclusions

Transportation agencies rely on freight demand forecasting models to develop, prioritize, and assess future infrastructure and policy scenarios. Advanced freight forecasting models that incorporate behavioral dimensions, including activity-based and truck touring models, allow for a wider range of policy evaluation and more detailed infrastructure planning. To date, such models have been hindered by a lack of relevant and available data. Fortunately, with new sources of big data evolving in the freight context, it is increasingly possible to practically implement advanced freight forecasting models. Unfortunately, the ubiquity of big data in and of itself does not close this critical data gap. This paper addresses the challenge of using big data for advanced freight travel demand modeling by developing and evaluating a method to extract representative and unique activity patterns from a common source of big data for trucks, e.g., passively collected GPS data.

A two-stage methodology is developed in which daily trip and stop characteristics are extracted from large streams of GPS pings (e.g., latitude, longitude, timestamp) and then used to find common but unique activity patterns defined as series of trips and stops. Heuristic based approaches to determine stop and trip characteristics were used in the first stage that fed into a K-means unsupervised clustering algorithm in the second stage. Using a statewide sample of GPS data for evaluation, we identified six activity patterns among 300,000 daily truck records. In relation to advanced freight models like ABMs, by reducing 300,000 daily truck activity patterns to a representative set of six, we aim to enable more efficient model calibration and validation.

About 32% of all trucks included in our statewide GPS sample belonged to the activity pattern cluster representing long-haul movements with a single stop, indicative of pass-through operations. The second most common patterns, approximately 50% in total if combined, captured medium-haul trips with several stops and a daily return to a home base but differed by the time of day in stop and trips took place. The least common pattern depicted short-haul trips with many stops connected by short trips, characteristics of local delivers or local mining operations.

Since truck GPS data used in our study was anonymous, it was not possible to directly ''observe" the demographic characteristics (e.g., industry-served or commodity carried) of the trucks within each representative pattern. Therefore, truck demographic characteristics associated with each activity pattern were inferred through visual comparisons of GPS trajectories and business and land use data. Representative activity patterns linked to industries can improve the ways in which the study extrapolates patterns derived from a sample to the population- a necessary step toward creating the data necessary for advanced freight forecasting models.

In future work, supervised machine learning can be used to predict commodity from operational features such as those described in this paper. For example, through supervised learning techniques, a predictive model can be trained to recognize the operational characteristics (e.g., daily activity patterns) that correspond to particular industries, given a large-enough sample of industry-labeled daily activity patterns. Further, while this study used only truck GPS data to distinguish activity patterns, addition of spatial data depicting business locations and/or land uses and the advent of spatial fusion approaches would allow us to identify the industry associated with each stop and relate it back to commodity specific activity patterns.

The developed model demonstrates that activity trajectories for a truck population can be approximated by a small set of representative patterns, containing some core trajectories, and that there are possible correlations among the demographics of commodities and the operational characteristics. In this way, we produce a novel dimension to passively collected mobile sensor data, that can be linked to the industry served and commodity carried without violating privacy concerns. Further, federal, state, and local transportation agencies can apply this approach to generate industry-specified activity pattern profiles that can be used for the development, calibration, and validation of advanced freight forecasting models. Ultimately, this approach can allow transportation agencies to satisfy the Moving Ahead for Progress in the 21st Century (MAP-21) and the Fixing America's Surface Transportation Act (FAST Act) goals by supporting the development of policy sensitive travel demand forecasting models.

Conflict of interest

All the authors have no conflict of interest with the funding entity and any organization mentioned in this article in the past three years that may have influenced the conduct of this research and the findings.

Acknowledgment

The authors thank the Arkansas Department of Transportation (ARDOT) for sponsoring the project that led to this paper. The authors declare that the contents of this article have not been published previously. All the authors have contributed to the work described, read, and approved the contents for publication in this journal.

References

Adler, T., Ben-Akiva, M., 1979. A theoretical and empirical model of trip chaining behavior. Transp. Res. Part B 13 (3), 243–257. https://doi.org/10.1016/0191-2615(79)90016-X.

Akter, T., Hernandez, S., Diaz, K.C., Ngo, C., 2018. Leveraging open-source GIS Tools to determine freight activity patterns from anonymous GPS data. Paper presented at the 2018 AASHTO GIS for Transportation Symposium.

Alho, A.R., Sakai, T., Chua, M.H., Jeong, K., Jing, P., Ben-Akiva, M., 2019. Exploring algorithms for revealing freight vehicle tours, tour-types, and tour-chain-types from GPS vehicle traces and stop activity data. J. Big Data Anal. Transp. 1 (2), 175–190.

Allahviranloo, M., Recker, W., 2013) Daily Activity Pattern Recognition by Using Support Vector Machines with Multiple Classes. doi: https://doi.org/10.1016/j.trb.2013.09.008.

Allahviranloo, M., Regue, R., Recker, W., 2017. Modeling the activity profiles of A Population. Transp. B: Transport Dyn. 5 (4), 426–449. https://doi.org/10.1080/21680566.2016.1241960.

Alpaydın, E. (2014). Introduction to Machine Learning (3. ed. ed.). Cambridge, Mass. [u.a.]: MIT Press. Retrieved from http://bvbr.bib-bvb.de:8991/F?func=service&doc_library=BVB01&local_base=BVB01&doc_number=027423356&sequence=000005&line_number=0001&func_code=DB_RECORDS&service_type=MEDIA.

Aziz, R., Kedia, M., Dan, S., Basu, S., Sarkar, S., Mitra, S., Mitra, P., 2016. Identifying and characterizing truck stops from GPS data. In Industrial Conference on Data Mining (pp. 168-182). Springer, Cham.

Bishop, C.M., 2016. Pattern Recognition and Machine Learning (softcover reprint of the original 1st edition. Springer, New York, NY. Retrieved from http://deposit.dnb.de/cgi-bin/dokserv?id=efb4c651772a4ab786a75b8fe8f0f0dd&prov=M&dok_var=1&dok_ext=htm.

Camargo, P., Hong, S., Livshits, V., 2017. Expanding the uses of truck GPS data in freight modeling and planning activities. Transp. Res. Rec. 2646 (1), 68–76. https://doi.org/10.3141/2646-08.

Chung, E., Shalaby, A., 2005. A trip reconstruction tool for GPS-based personal travel surveys. Transp. Plann. Technol. 28 (5), 381–401. https://doi.org/10.1080/03081060500322599.

Corro, K.D., Akter, T., Hernandez, S., 2019. Comparison of overnight truck parking counts with GPS-derived counts for truck parking facility utilization analysis. Transp. Res. Rec. 2673 (8), 377–387. https://doi.org/10.1177/0361198119843851.

Fhwa, 2001. Analysis of the Vehicle Inventory and Use Survey for Trucks with Five Axles or More Retrieved from https://www.fhwa.dot.gov/reports/tswstudy/vius97.pdf.

Fhwa, 2014. All Road Network of Linear Referenced Data (ARNOLD) Reference Manual Retrieved from https://www.fhwa.dot.gov/policyinformation/hpms/documents/arnold_reference_manual_2014.pdf.

Fhwa, 2018. Status of The Nation's Highways, Bridges, And Transit Conditions and Performance: 23rd Edition: Part III: Highway Freight Transportation - Report to Congress Retrieved from https://ops.fhwa.dot.gov/freight/infrastructure/nfn/rptc/cp23hwyfreight/iii_ch11.htm.

Fmcsa, 2020. Summary of Hours of Service Regulations Retrieved from https://www.fmcsa.dot.gov/regulations/hours-service/summary-hours-service-regulations.

Gingerich, K., Maoh, H., Anderson, W., 2016. Classifying the purpose of stopped truck events: An application of entropy to GPS data. Transportation Research Part C: Emerging Technologies 64, 17–27.

Giovannini, L., 2011. A Novel Map-Matching Procedure for Low-Sampling GPS Data with Applications to Traffic Flow Analysis. https://doi.org/10.6092/unibo/amsdottorato/3898. Retrieved from https://www.openaire.eu/search/publication?articleId=od______1754::2e76bee797112fda11280f4851def321.

Hunt, J.D., Stefan, K.J., 2007. Tour-based microsimulation of urban commercial movements. Transp. Res. Part B 41 (9), 981–1013.

Jiang, S., Ferreira, J., González, M., 2012. Clustering Daily Patterns of Human Activities in The City. Data Min. Knowl. Disc. 25 (3), 478–510. https://doi.org/10.1007/s10618-012-0264-z.

Jing, P., 2018. Identifying and Modeling Urban Truck Daily Tour-Chaining Patterns. Massachusetts Institute of Technology). Doctoral dissertation.

Joubert, J.W., Meintjes, S., 2015. Repeatability & reproducibility: Implications of using GPS data for freight activity chains. Transp. Res. Part B 76, 81–92.

Kodinariya, T.M., Makwana, P.R., 2013. Review on Determining Number of Cluster in K-Means Clustering. International Journal 1 (6), 90–95.

Kuppam, A., Lemp, J., Beagan, D., Livshits, V., Vallabhaneni, L., Nippani, S., 2014. Development of A Tour-Based Truck Travel Demand Model Using Truck GPS Data. Paper presented at the 93rd Annual Meeting of the Transportation Research Board.

Laranjeiro, P.F., Merchán, D., Godoy, L.A., Giannotti, M., Yoshizaki, H.T., Winkenbach, M., Cunha, C.B., 2019. Using GPS data to explore speed patterns and temporal fluctuations in urban logistics: The case of São Paulo, Brazil. J. Transp. Geogr. 76, 114–129.

Li, S., Lee, D., 2017. Learning Daily Activity Patterns with Probabilistic Grammars. Transportation 44 (1), 49–68. https://doi.org/10.1007/s11116-015-9622-1.

Liao, C., 2009. Using Archived Truck GPS Data for Freight Performance Analysis On I-94/I-90 From the Twin Cities to Chicago. University of Minnesota Center for Transportation Studies, Retrieved from. https://conservancy.umn.edu/handle/11299/97668.

Liu, F., Janssens, D., Cui, J., Wang, Y., Wets, G., Cools, M., 2014. Building A Validation Measure for Activity-Based Transportation Models Based on Mobile Phone Data. Expert Syst. Appl. 41 (14), 6174–6189. https://doi.org/10.1016/j.eswa.2014.03.054.

Liu, F., Janssens, D., Cui, J., Wets, G., Cools, M., 2015. Characterizing Activity Sequences Using Profile Hidden Markov Models. Expert Syst. Appl. 42 (13), 5705–5722. https://doi.org/10.1016/j.eswa.2015.02.057.

Ma, X., McCormack, E.D., Wang, Y., 2011. Processing Commercial Global Positioning System Data to Develop A Web-Based Truck Performance Measures Program. Transp. Res. Rec. 2246 (1), 92–100. https://doi.org/10.3141/2246-12.

Ma, X., Wang, Y., McCormack, E., Wang, Y., 2016. Understanding freight trip-chaining behavior using a spatial data-mining approach with GPS data. Transp. Res. Rec. 2596 (1), 44–54.

Nepal, S. A., Farnsworth, S. P., & Pearson, D. F. (2006). San Antonio Area Commercial Vehicle Survey Technical Summary. A report prepared by the Texas Transportation Institute for the Texas Department of Transportation Travel Survey Program.

Nepal, S., Farnsworth, S., & Pearson, D. (2005). Amarillo Area Commercial Vehicle Survey Technical Summary. A report prepared by the Texas Transportation Institute for the Texas Department of Transportation Travel Survey Program.

Ng, A., 2012. Clustering with the k-means algorithm. Machine Learning.

Oka, H., Hagino, Y., Kenmochi, T., Tani, R., Nishi, R., Endo, K., Fukuda, D., 2019. 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. Transp. Res. Part E 129, 305–324.

Pline, J.L., 1999. Traffic engineering handbook. Inst. of Transportation Engineers, Washington, DC.

Quddus, M., Washington, S., 2015. Shortest Path and Vehicle Trajectory Aided Map-Matching for Low Frequency GPS Data. Transportation Research Part C: Emerging Technologies 55, 328–339. https://doi.org/10.1016/j.trc.2015.02.017.

Ruan, M., Lin, J., Kawamura, K., 2012. Modeling urban commercial vehicle daily tour chaining. Transp. Res. Part E 48 (6), 1169–1184. https://doi.org/10.1016/j.tre.2012.06.003.

Sánchez-Díaz, I., Holguín-Veras, J., Ban, X.J., 2015. A time-dependent freight tour synthesis model. Transp. Res. Part B 78, 144–168.

Sharman, B.W., Roorda, M.J., 2013. Multilevel modelling of commercial vehicle inter-arrival duration using GPS data. Transp. Res. Part E 56, 94–107.

Shoval, N., Isaacson, M., 2007. Tracking tourists in the digital age. Ann. Tourism Res. 34 (1), 141–159. https://doi.org/10.1016/j.annals.2006.07.007.

Thakur, A., Pinjari, A.R., Zanjani, A.B., Short, J., Mysore, V., Tabatabaee, S.F., 2015. Development of algorithms to convert large streams of truck GPS data into truck trips. Transp. Res. Rec. 2529 (1), 66–73. https://doi.org/10.3141/2529-07.

Waldroup, P.W., Hellwig, H.M., 2000. The potential value of morning and afternoon feeds for laying hens. J. Appl. Poultry Res. 9 (1), 98–110.

Yang, X., Sun, Z., Ban, X.J., Holguín-Veras, J., 2014. Urban freight delivery stop identification with GPS data. Transp. Res. Rec. 2411 (1), 55–61.

Yang, Y., Yao, E., Yue, H., Liu, Y., 2010. Trip chain's activity type recognition based on support vector machine. J. Transp. Syst. Eng. Inf. Technol. 10 (6), 70–75. https://doi.org/10.1016/S1570-6672(09)60073-8.

You, S.I., Ritchie, S.G., 2018. A GPS data processing framework for analysis of drayage truck tours. KSCE J. Civil Eng. 22 (4), 1454–1465.

You, S.I., Chow, J.Y., Ritchie, S.G., 2016. Inverse vehicle routing for activity-based urban freight forecast modeling and city logistics. Transportmet. A: Transp. Sci. 12 (7), 650–673.

Zanjani, A.B., Pinjari, A.R., Kamali, M., Thakur, A., Short, J., Mysore, V., Tabatabaee, S.F., 2015. Estimation of statewide origin-destination truck flows from large streams of GPS data. Transp. Res. Rec. 2494 (1), 87–96. https://doi.org/10.3141/2494-10.