Freight activity-travel pattern generation (FAPG) as an enhancement of freight (trip) generation modelling
Methodology and case study

Pani, Agnivesh; Sahu, Prasanta K.; Tavasszy, Lóránt; Mishra, Sabya

DOI
10.1016/j.tranpol.2023.09.020

Publication date
2023

Document Version
Final published version

Published in
Transport Policy

Transport Policy 144 (2023) 34–48

Contents lists available at ScienceDirect

Transport Policy

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

Freight activity-travel pattern generation (FAPG) as an enhancement of freight (trip) generation modelling: Methodology and case study

Agnivesh Pani [a], Prasanta K. Sahu [b,*], Lóránt Tavasszy [c], Sabya Mishra [d]

[a] Department of Civil Engineering, Indian Institute of Technology (BHU) Varanasi, Varanasi, Uttar Pradesh, 221005, India
[b] Department of Civil Engineering, Birla Institute of Technology and Science Pilani, Hyderabad, Telangana, 500078, India
[c] Faculty of Technology, Policy and Management (TPM), Delft University of Technology, PO Box 5015, 2600, GA, Delft, the Netherlands
[d] Department of Civil Engineering, University of Memphis, Memphis, TN, 38152, USA

ARTICLE INFO

Keywords:
Freight activity-travel patterns
Shipper behavior
Activity-based modelling
Segmentation analysis

ABSTRACT

Trip-based models and activity-based models represent two extreme ends of the spectrum of travel demand models in data granularity requirement and ability to reflect the underlying motivation to travel. Modelling of representative freight activity-travel patterns (RFAPs) has the potential to serve as the bridge between these approaches. RFAP clusters represent homogeneous groups of establishments, where utility maximization models predict the probability that an establishment belongs to a particular cluster. However, it is still an open question how to define, interpret and model activity-travel patterns in the context of freight system. To answer this question, this study conducted a large-scale establishment-based freight survey (EBFS) in seven cities of India and resulted in a sample of 432 establishments and their 1613 shipment records. In the first part, this paper proposes a novel approach for identifying RFAPs based on the notion that "activities" that inspire trip-making for passenger is equivalent to "freight orders" in the case of establishments. The cluster analyses revealed the presence of three well separated main clusters and nine less separated nested clusters. Through interpretation and labelling of these RFAPs, freight travel market is categorized into useful segments. The results suggested that *a priori* industrial classification systems used in trip-based models are overly simplified representations of the complex structure of the travel patterns. In the last part, freight activity-travel pattern generation (FAPG) models are developed which predicts the probability that an establishment exhibits a particular RFAP. The FAPG models developed using these RFAPs could replace the traditional freight generation (FG) and freight trip generation (FTG) models due to its ability to convert the assigned activity-patterns to trips or tonnage. For example, FAPG model suggest that at an employment level of 120, there is a 56% probability that establishments will exhibit MDV-HFMH (medium duty vehicles - high frequency medium haul) pattern which, in turn, implies that FP = 1630 tons/year; shipment frequency, i.e., FTP = 8 trips/week; length of haul = 240.6 km and commercial vehicle type choice = MDV. Thus, FAPG models can present an enhanced representation of freight flows since both FG and FTG are jointly modeled in this approach. That is, the best features of both commodity-based modelling (i.e., ability to capture the fundamental mechanism that drives freight demand) and vehicle-based modelling (i.e., ability to capture freight traffic implications) are included in FAPG models. The study findings are expected to assist in identifying the variations in establishments' preferences so that it is possible to identify the type of transport supply improvements that the establishments will respond to accurately, and thus prioritize the infrastructure investments. Moreover, the discussions on these findings are expected to improve the behavioral and spatial foundations of traditional freight models.

1. Introduction

It is imperative that the complex travel behavior patterns need to be understandable and predictable so that facilities can be planned and designed to cater to the different needs of a diverse population (Cheng et al., 2019; Li and Tong, 2016). The classification of complex travel

* Corresponding author.
E-mail addresses: agniveshpani@gmail.com (A. Pani), prasantsahu222@gmail.com (P.K. Sahu), L.A.Tavasszy@tudelft.nl (L. Tavasszy), smishra3@memphis.edu (S. Mishra).

https://doi.org/10.1016/j.tranpol.2023.09.020
Received 1 June 2023; Received in revised form 15 September 2023; Accepted 24 September 2023
Available online 5 October 2023

patterns into homogeneous groups, in and of itself, is widely acknowledged to be the first step towards understanding individual travel behavior (Joh et al., 2001; Yasmin et al., 2017). The search for identification of these homogeneous groups is motivated by the need for estimating disaggregate-level travel demand models, which requires some level of simplification of underlying complexities of behavioral patterns to estimate feasible models (Dharmowijoyo et al., 2017; Kulkarni and McNally, 2001). From a theoretical perspective, identification of such groups help to describe the determinants and constraints for travel in various travel segments (Kroesen, 2014). From an applied perspective, grouped travel demand predictions for such groups assist transport authorities and operators in understanding the transportation needs and service requirements of population subgroups (Bieger and Laesser, 2002; Gunay et al., 2016). The activity-based modelling (ABM) approach extends this concept into grouping similar activity-travel patterns since travel behavior is directly connected to the need for participating in activities. These groups, often referred as representative activity-travel patterns (RAPs) form the fundamental unit of analysis in the case of ABM approach, whereas the conventional four-step modelling approach (FSM) considers aggregate or zonal trips as the unit of analysis. Therefore, ABM approach explicitly supersedes the FSM approach due to its ability to reflect underlying travel behavior and, in turn, to be responsive to novel policies towards demand management and infrastructure augmentation (McNally, 1996). Despite the existence of a body of systematic scientific evidence on grouping activity-travel patterns (Diana and Mokhtarian, 2009; Hanson and Huff, 1986; Molin et al., 2016) and activity-based travel demand analysis of individuals or households (Bowman and Ben-Akiva, 2000; Pinjari and Bhat, 2011), the research on grouping freight activity-travel patterns or activity-based freight transport modelling of establishments have been modest at best, possibly be due to the difficulties in obtaining the requisite freight data. Thus, the state-of-practice in modelling freight transport is yet limited to FSM approach (Samimi et al., 2014) which concentrates on aggregate freight flows (in number of trips or quantity of tonnage) instead of individual logistical decisions by establishments.

The classification of establishments is a crucial step in FSM approach for improving the prediction accuracy of model estimates and organizing the models in such a way that they complement the forecast requirements of land-use ordinances and of policy interventions (Gunay et al., 2016; Tavasszy and De Jong, 2014). Lately, there is a growing interest in identifying the most suitable classification system that leads to homogeneous classes of establishments which are internally consistent in terms of their demand and trip characteristics (Campbell et al., 2012; Chandra et al., 2021a; Gonzalez-Feliu and Sánchez-Díaz, 2019; Holguín-Veras et al., 2012; Pani et al., 2019b; Pani and Sahu, 2019a). An examination of literature reveal that the previous studies have predominantly used *a priori* segmentation approach, rather than the data-driven *a posteriori* segmentation approach. The *a priori* approach assumes that there is some knowledge about the homogeneous travel segments in advance and they are distinguishable by existing classification systems. One of the reasons for this rather reductive assumption is that it conceals the limitation of FSM approach in distinguishing the diverse population of establishments. The *a priori* approach used by most of the previous studies does not enable us to know whether the existing classification systems are indeed the most important determinants of travel patterns. Also, among the several *a priori* classification systems, it is not yet known whether the groups segmented on the basis of land-use or industrial classification truly represent the minimum variability groups with respect to observed freight travel patterns. Finally, and perhaps more importantly, the existence such orderly, identifiable and isomorphic freight travel patterns among *a priori* segments, although an attractive and popular notion, is not yet empirically tested in the literature. Therefore, the aim of this research is to identify and propose a typology of representative freight activity-travel patterns (RFAPs) of establishments, analogous to RAPs of households. The focus is to examine how freight trips are spatially distributed and how do they vary depending upon the shipment sizes and frequency. More specifically, how often do freight trips occur, ship a specific quantity and to what degree does distance determine the shipment size and frequency? To what extent are the total tonnage shipped by an establishment influenced by the distance covered in freight trips? So, by exploring the geographies of freight activity-travel patterns, the objective is to understand not only how goods are moved but also the quantity of goods moved and how far or frequently they are transported in modern supply-chains. The obtained RFAPs are further used to develop a probabilistic model that is able to allocate establishment to a category in the proposed typology – the resulting allocation model is termed as freight activity-travel pattern generation (FAPG). The resulting prediction from FAPG models are expected to provide an enhanced representation of freight flows since both FG and FTG components are embedded in the model. It may also be noted that the analyses in this paper are developed using conventional trip-based establishment freight survey (EBFS) data. This research thus represents a test of the suitability of conventional EBFS data as a proxy for the rarely available activity-based freight data in construction of FAPG models. In short, it is expected that this study will open the possibility for planners to build improved travel demand models in cases which activity-based survey cannot be undertaken considering the high project expenses. This paper is also expected to demonstrate the feasibility and practicality of activity-based freight transport modelling in developing countries like India.

This paper is structured in seven sections out of which this is the first. The following section presents a brief background of activity-based freight transport modelling and defines activity-travel patterns in the context of freight movements. Section 3 illustrates the research design and data used in this study along with all-encompassing aspects of data collection framework, methodological approach and analysis methods. The RFAPs are identified, labeled and interpreted in Section 4. The model estimation results and interpretations of FAPG models are given in Section 5. The last two sections contain the research implications and conclusions.

2. Research background and rationale

2.1. Activity-based modelling approach

The fundamental difference between trip-based FSM approach and activity-based ABM approach is that the former directly focuses on “trips” without recognizing the motivation or reason that necessitates trip-making - i.e., activity participation (Pinjari and Bhat, 2011). In doing so, trip-based FSM approach tend to ignore the diversity among individuals and constructs models for aggregate (zonal) trips in four well known steps - trip generation, trip distribution, mode split and route assignment (Chandra et al., 2021b; Pani et al., 2019a; Pinjari and Bhat, 2011). While the earliest trip-based models quantify the number of trips generated from a zone as a function of zonal characteristics, the advances in modelling techniques necessitated a paradigm shift to disaggregate-level trip-based models (Ortúzar and Willumsen, 2011). These models view individuals (household or establishment) as the decision-making unit and quantify the trips made by individuals between the zones in the study area. Despite the paradigm shift to disaggregate-level models, trip-based FSM approach continue to exhibit several limitations out of which few are as follows (i) inability to consider trip linkages; (ii) far-fetched assumption of independence between decisions involved in four steps of the modelling process and (iii) aggregation bias due to the assumption that group characteristics (e.g., household type, income level, industry sector, land-use type) are shared by all the individuals who are members of a particular *a priori* segments (Castiglione et al., 2014). The first limitation is largely overcome with the advent of tour-based FSM approach in which individual travel is divided into trip chains beginning and ending at home or work. Subjected to the availability of trip chaining information from individuals, few of the prevalent travel demand models use tour-based FSM approach, including freight demand models. Calgary commercial vehicle movement model (Hunt and Stefan, 2007) is the first large-scale tour based freight model reported in literature. In the Calgary study, the authors developed an agent-based microsimulation approach to model tour generation, duration, number of stops, and destinations. The freight trip chaining behavior is further investigated by analytical methods such as entropy maximization (Wang and Holguín-Veras, 2008), time dependent freight tour synthesis models (Sánchez-Díaz et al., 2015) and spatial price equilibrium models (Holguín-Veras et al., 2016b). The tour-based FSM approach, although popular in practice, is still rather limited due to the lack of behavioral foundation, a trait jointly shared by trip-based FSM approach as well. The solution to the limitations of the trip-based and tour-based FSM approach is the promising alternative of activity-based modelling (ABM) approach which acknowledge the fact that travel needs are byproducts of the necessity to partake in activities spread out over space and time. These models are based on behavioral theories about how decisions are made regarding where to participate in activities, when to participate and how to get to these activities. Consequently, an individual's activity patterns reflect the individual's travel patterns. Since the modelling objects are decisions, activity-based models are deemed to be significantly more useful in quantifying how investments, policies, or other changes affect the individual travel behavior (Samimi et al., 2014).

2.2. Defining activity-travel patterns

A widely recognized approach to overcome data insufficiency for activity-based modelling is to consider "activity-travel patterns" as the fundamental unit of analysis (Kulkarni and McNally, 2001). These activity travel-patterns comprises of several smaller, inter-connected activity-travel decisions on activity type, frequencies, schedule, sequencing and distances (Buliung and Kanaroglou, 2006; Castiglione et al., 2014; Mitra and Buliung, 2012; Yasmin et al., 2017). Given the inter-dependence of these decisions, activity-travel patterns are typically assumed to be indistinguishable and therefore often treated as a joint set options for the individual (Bhat and Singh, 2000; Recker et al., 1986a). The systematic body of literature in this research area (Kulkarni and McNally, 2001; McNally and Rindt, 2007; Recker et al., 1986b) suggest that various classification techniques can be used for identifying homogeneous and distinct groups of "representative activity-travel patterns" (RAPs). More recently, Dashtestaninejad et al. (2014) adopted this approach to identify distinct groups rural residents based on their activity-travel patterns. The identified RAPs are used to develop activity-travel pattern generation models on the theoretical premise of utility maximization (Kulkarni and McNally, 2001). Further on, there has been three significant improvements over the basic concept of RAPs by incorporating (i) daily activity schedules (Bowman and Ben-Akiva, 2000); (ii) tour choices (Wen and Koppelman, 2000); (iii) socio-demographic and transportation level of service attributes (Bhat et al., 2004). The daily activity schedule model or commonly termed as "BB system" (Bowman and Ben-Akiva, 2000), for example, is a nested logit model for a choice set of 55 alternatives based on activity-travel patterns and a system of conditional tours. In BB system, activity-travel patterns are defined in terms of a primary activity (home, work, school or other), a primary tour type (e.g., home to work and back) and the number and purpose of secondary tours. The tour-choice modelling system (Wen and Koppelman, 2000) is largely similar to the BB system, except that it excludes mode and destination choice. The third, and perhaps the most comprehensive activity-based model system labeled as CEMDEP (Bhat et al., 2004) differentiates between workers and non-workers in the population and simulates their activity-travel patterns in three layers: patterns (activity type, transport mode, number of stops, commute duration), tours (activity type, transport mode, number of stops, tour duration) and stops (activity type, stop duration, travel time to stop and stop location). Overall, the models discussed in this section cover the state-of-the-art of activity-based travel demand models based on utility maximization. The other prominent alternatives in developing activity-based models recorded in literature (constraint-based models, computational process models, agent-based micro-simulation models) are not relevant to the scope of this study and, therefore, not described in this paper.

2.3. Activity-based freight transport modelling

The trip-based FSM approach is still the state-of-the-practice in modelling freight transportation due to the difficulties associated with collecting freight information (Chow et al., 2010). This approach starts with collecting freight generation (FG) and freight trip generation (FTG) data for developing quantitative models. FG focuses on the quantity of tonnage transported in trips, whereas FTG refers to the number of truck trips (Alho and de Abreu e Silva, 2016; Pani et al., 2018). Apart from generating the trip ends or tonnage ends required as inputs for subsequent stages in FSM approach, both FG and FTG quantifications are necessary to quantify the traffic impacts and freight needs of establishments in urban areas (Sakai et al., 2018; Sánchez-Díaz et al., 2016). Notwithstanding, there is a pressing need for developing innovative freight models that capture the underlying decision-making process that necessitate freight movements is reported in the literature (Chow et al., 2010; Hensher and Figliozzi, 2007; Pani et al., 2019a; Sánchez-Díaz et al., 2015). The ABM approach in passenger modelling regards activity-travel pattern as the central modelling object, which is defined as the revealed pattern of behavior represented by travel and activities of passengers over a specified time period (Recker et al., 1986a). Of course, this cannot be the central modelling object in the case of activity-based freight transport modelling. In order to adapt ABM approach to freight modelling, freight orders are considered to be the freight system equivalent to activities in passenger transportation (Chow et al., 2010; Liedtke and Friedrich, 2012). The fundamental premise of this analogy is that freight travel demand is derived from establishments' need to fulfill "freight orders" which, in turn, is the logical counterpart of "activities" driving passenger travel demand. While the semantical appropriateness of analogizing shipments generated by freight orders to trips generated by activities (e.g., shopping, work) is debatable, this usage is logically reasonable as both terms fundamentally represent the underlying motivation for travel. Thus, the conceptual underpinning for activity-based freight transport modelling is that the logistical behavior establishments can be captured if freight orders are considered to be the central modelling object (Liedtke and Schepperle, 2004). These models can be developed using utility-maximization approach to represent how establishments make decisions about freight orders in the presence of constraints, including the decisions about where to forward the shipments, which shipment size to be selected and for what frequency the shipments are forwarded in a standard time frame like a week.

2.4. Defining freight activity-travel patterns

Despite having a systematic body of literature on activity-based passenger models that use RAPs as the unit of analysis, the concept of representative freight activity-travel patterns (RFAPs) is not yet explored for freight transport modelling. The underlying hypothesis of this approach is that there exist groups of individuals with similar travel behaviors and that by distinguishing these groups it is possible to develop activity-travel pattern generation models (Kulkarni and McNally, 2001). According to Liedtke and Schepperle (2004), freight activity-travel pattern may be defined using attributes of freight orders so that it explains how the logistical decisions are undertaken. Such attributes reported in literature may be categorized into three groups: (i) physical expression of total economic exchanges between stakeholders, such as freight generation (Holguín-Veras et al., 2011; Sahu and Pani, 2020); (ii) commercial vehicle choice determinants such as shipment size (Holgun-Veras, 2002), (iii) variables denoting transportation supply characteristics such as shipment frequency, transportation cost, transit time, length of haul and damage rates (Jiang et al., 1999). However, many of these attributes are not easily measurable since freight transportation system consists of multiple agents and numerous inter-connected activities. These interacting agents in freight system are of three broad types: shippers, carriers and receivers (Holguín-Veras et al., 2012). Shippers are agents which produce or ship freight, such as raw material production sites, distribution or assembling sites, manufacturing units and wholesale retailers. Carriers are agents hired by shippers for transporting freight from one end to another which includes private carriers, for-hire carriers, third party logistics providers and freight forwarders. Receivers are agents to whom the shipment is destined, including wholesale traders, retailers, intermediate consumers and end-consumers. Amongst these three groups in the freight system, shippers are reported to have the maximum information regarding freight orders (Holguín-Veras et al., 2012). This implies that the attributes of freight orders are best suited to be collected by establishment-based freight surveys (EBFS) targeted at shippers in an urban area.

To conclude, the research background can be recapitulated into four fronts. First, there is a concerning absence of practice-oriented activity-based freight transport models which reflect the underlying logistical decisions. Second, freight orders are the logical counterpart of the term activities defined in the case of passengers since both represent the causal motivation for trip-making. Third, homogeneous segments of freight activity-travel patterns, if properly distinct, could be used as the fundamental unit of analysis for developing activity-based freight transport models. Fourth, establishment-based freight surveys targeted at shippers are best suited for collecting attributes of freight orders which can be used for grouping freight activity-travel patterns. Based on these insights, this paper attempts to improve the knowledge on freight activity-travel patterns through a data collection exercise, freight travel segmentation and a utility maximization modelling approach.

3. Research design

The research methodology and its distinction with traditional industry-aggregation approach for model estimation is presented in Fig. 1. As shown, the traditional aggregation approach involves the usage of industrial classes, as given in NCFRP-37 (Holguín-Veras et al., 2016a), land use classes (Holguín-Veras et al., 2012) or ensembles of '*a posteriori*' segments based on '*a priori*' classes, such as industry clusters (Gunay et al., 2016; Pani and Sahu, 2019b). The industrial classes, such as NAICS, NACE, ISIC and the land-use classes such as CPA, NAPCS, CPC are widely used in many of the traditional FG/FTG models in the literature (Pani et al., 2022).

In the first stage of analysis, the RFAP typology proposed in this paper is estimated based on annual freight production (tons), average shipment size (tons), shipment frequency (trips/week) and average length of haul (km). The optimal number of RFAPs of establishments is determined by Gap statistics method which formalizes the heuristics associated with locating elbow criterion in intra-cluster variation plots. The extracted travel segments are labeled and interpreted in terms of observable traits and industry sector profiles. The significance of extracted clusters is that they can be considered as a proxy of activity-travel decisions based on freight orders. In the second stage of analysis, the freight activity-travel pattern generation (FAPG) models are developed subsequently for allocating establishments into their travel segments using business size indicators, commodity characteristics and locational characteristics. The data and analysis methods used in this study are explained below in sequence.

3.1. Data

The research is undertaken in Kerala, a strategically important State in west-coast trade corridor of India due to the presence of Cochin

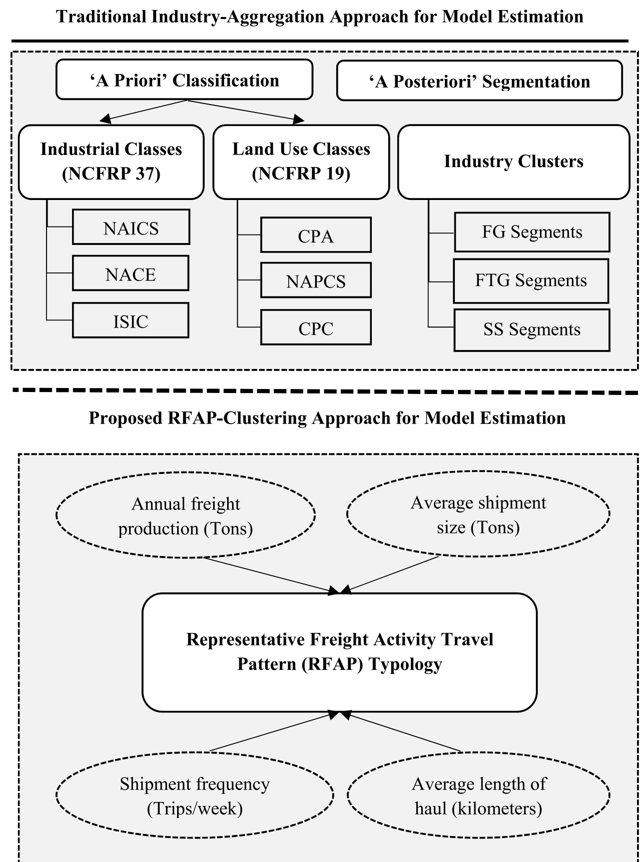

Fig. 1. Proposed Methodological Approach involving Representative Freight Activity Travel Patterns (RFAP).

seaport which is one of the major gateways to international shipping traffic. Due to this trade corridor, most of the industries in Kerala developed in the vicinity of the coastal belt and industrial growth has been along the spine of National Highway (NH) network since land is constrained by the sea at south and west. The study cities are: Cochin, Calicut, Malappuram, Kannur, Palakkad, Thrissur and Kottayam. In order to provide a quantitative evaluation of freight demand characteristics, EBFS was conducted during the year 2016 in the study cities. The scope of the survey included shippers (industrial establishments) in manufacturing, wholesalers, and raw material production sites in Kerala. The questions included in the survey instrument can be divided into three types: (i) establishment characteristics; (ii) physical characteristics of the product or commodity; (iii) spatial and flow characteristics of shipments. Establishment characteristics include the establishment type (ISIC category), location (geo-code) and different measures of business size, such as number of employees, gross floor area and number of years in business. The next series of questions collected information on the attributes of the goods to be transported, such as weight and value. Spatial and flow characteristics of freight shipments included in the questionnaire were shipment size, frequency, length of haul, origin and destination of the shipments. After data cleaning, the final sample consisted of 432 establishments providing information about 1613 daily shipments. More information on the data collection framework, survey coverage, sampling strategies and response rates can be found in Pani and Sahu (2019b). ISIC system was used for segregating the cleaned data sample into homogeneous groups of economic activities. The establishment groups are as follows: (1) ISIC 10: Food products, (2) ISIC 11: Beverages, tobacco and related products, (3) ISIC 13: Textile mills, (4) ISIC 14: Textile products, (5) ISIC 16: Wood, wood products, furniture and fixtures, (6) ISIC 17–18: Paper, paper products and printing, (7) ISIC 20–21: Basic chemicals, chemical products and pharmaceuticals, (8) ISIC 22: Plastic and rubber products, (9) ISIC 23: Non-metallic mineral products, (10) ISIC 24–25: Basic metal, alloy, metal products, (11) ISIC 26–28: Machinery and equipment, (12) ISIC 29–30: Transportation equipment and (13) ISIC 32: Other manufacturing industries. The stratifications of the collected sample with respect to study cities and industry sectors are given in Table 1. More information about the classification system can be found in Pani and Sahu (2019a).

As shown, the proportion of sampled establishments for study cities (part I) or industrial sectors (part II) are comparable to the proportion of total establishments in the entire population. The average deviation of establishment proportion between sample and population are found to be 3.02% across industry sectors and 4.4% across study cities. Additionally, it can be seen that the average sample employment and average population employment are reasonably close to each other except in the case of Malappuram city. Thus, the comparison of sample and population characteristics clearly underlines the statistical validity of sampling design and provides a strong indication of the representativeness of the study sample.

3.2. Analysis methods

3.2.1. Cluster analysis

The variables related freight activity-travel pattern are used to classify the establishments into distinct clusters. K-means clustering algorithm (KMCA) is used to identify these clusters in the data. This algorithm partitions the data and arrive at a cluster solution such that the within-cluster variation is minimized (Everitt et al., 2011). Although efficient, one of the shortcomings of KMCA is that it assumes prior knowledge of the data in order to choose appropriate number of clusters. The optimal clustering, in turn, depends largely on the selection of appropriate number of clusters. Specifying too many clusters produce very small clusters that are difficult to define, while specifying too few clusters makes it impossible to differentiate between important factors as clusters are inordinately large. This paper thus uses two complementary methods to arrive at the most parsimonious cluster solution: (a) Direct or "Elbow" method: considers the total within cluster sum of squares (W_k) as a function of number of clusters (k) and chooses optimum number of clusters ($\widehat{k}$) as the cluster solution in which adding another cluster doesn't contribute to significant reduction of W_k; (b) Gap statistics method: consists of computing a test statistic value which compares the total within cluster sum of squares of observed data with that of a null reference distribution (Tibshirani et al., 2001). The mathematical formulation of these two methods is explained as follows. Consider a dataset $D = \{x_{ij}\}\ \forall i = 1, 2\ldots n$ and $j = 1, 2\ldots p$, which consists of p features (i.e., clustering variables) measured on n independent observations. Let $dist(i, i')$ denote the distance between observations i and i', typically computed as the Euclidean distance $\sum_j (x_{ij} - x_{i'j})^2$. Let the k clusters are denoted as $C_1, C_2\ldots C_r\ldots C_k$ and the centroids of observations in these clusters are signified as $\mu_1, \mu_2\ldots\mu_r\ldots\mu_k$ respectively. Then, the pooled within-cluster sum of squares (W_k) is computed as given in Eq. (1). It is worth mentioning that W_k measures the compactness of clustering. The direct method of determining optimum number of clusters is based on the location of "elbow criterion" in the plot between W_k and k. The elbow criterion looks for the smallest value of k such that the incremental reduction in W_k with the addition of a cluster is diminished. The elbow criterion, however, often tends to be ambiguous and necessitate subjective judgements by the analyst. This shortcoming is overcome by the second method which provides a statistical procedure that formalizes the heuristics associated with elbow criterion. 1

$$\text{Within Cluster Sum of Squares } (W_k) : W_K = \sum_{r=1}^{k} \sum_{i \in C_r} dist(i, \mu_r) \tag{1}$$

$$\text{Gap Statistic } (G_k) : G_K = E_n^*\{\log W_k\} - log\ W_k \tag{2}$$

$$: G_K = \frac{1}{B} \sum_{b=1}^{B} \log\ (W_{kb}^*) - log\ (W_k) \tag{3}$$

$$\text{Mean of } W_{kb}^* : \overline{w} = \frac{\sum_{b=1}^{B} \log\ (W_{kb}^*)}{B} \tag{4}$$

$$\text{Standard Deviation of } W_{kb}^* : sd(k) = \sqrt{\frac{\sum_{b=1}^{B} \{\log\ (W_{kb}^* - \overline{w})\}^2}{B}} \tag{5}$$

$$\text{Standard Error } (s_k) : s_k = \sqrt{1 + \frac{1}{B}} * sd(k) \tag{6}$$

$$\text{Optimal Number of Clusters } (\widehat{k}) : \widehat{k} = \text{Smallest } k \text{ such that } G_k \geq G_{k+1} - s_{k+1} \tag{7}$$

The gap statistic method standardizes the comparison of W_k by contrasting $log\ (W_k)$ of observed data with a null reference distribution of the data, i.e., a distribution with no apparent clustering structure. The estimate of optimal number of clusters ($\widehat{k}$) in this method is the value for which $log\ (W_k)$ falls the farthest below the $log\ (W_k^*)$ of reference distribution. In line with this, gap statistics (G_k) values are formulated as

Table 1
Characteristics of the survey sample compared to the establishment population in Kerala.

Part I: Sample Stratification ~ Study Cities					
City	EBFS Sample	Population	City	EBFS Sample	Population
Number of Establishments *(Percentage)*			Average Employment per Establishment		
Calicut	91 *(21.1%)*	6900 *(12.7%)*	Calicut	23	24
Cochin	75 *(17.4%)*	13,045 *(24.1%)*	Cochin	25	26
Malappuram	61 *(14.1%)*	11,984 *(22.1%)*	Malappuram	6	15
Kannur	53 *(12.3%)*	4761 *(8.8%)*	Kannur	18	22
Palakkad	57 *(13.2%)*	5330 *(9.8%)*	Palakkad	31	35
Thrissur	58 *(13.4%)*	7167 *(13.2%)*	Thrissur	37	40
Kottayam	37 *(8.6%)*	4983 *(9.2%)*	Kottayam	22	28
Total	432	54,170	Combined	23	27
Part II: Sample Stratification ~ Industry Sectors					
Industry Sector	EBFS Sample	Population	Industry Sector	EBFS Sample	Population
Number of Establishments *(Percentage)*			Number of Establishments *(Percentage)*		
ISIC 10	98 *(22.7%)*	14,176 *(26.2%)*	ISIC 22	37 *(8.6%)*	5555 *(10.3%)*
ISIC 11	22 *(5.1%)*	1546 *(2.9%)*	ISIC 23	27 (6.3%)	1497 *(2.8%)*
ISIC 13	11 *(2.5%)*	3003 *(5.5%)*	ISIC 24-25	51 *(11.8%)*	2846 *(5.3%)*
ISIC 14	18 *(4.2%)*	5665 *(10.5%)*	ISIC 26-28	37 *(8.6%)*	6415 *(11.8%)*
ISIC 16	50 *(11.6%)*	5942 *(11.0%)*	ISIC 29-30	10 *(2.3%)*	1352 *(2.5%)*
ISIC 17-18	14 *(3.2%)*	1142 *(2.1%)*	ISIC 32	13 *(3.0%)*	2618 *(4.8%)*
ISIC 20-21	44 *(10.2%)*	2413 *(4.5%)*	Total	432	54,170

[a] Information sources for establishment population of Kerala as follows. 1. Number of establishments: Economic census of India (Central Statistics Office, 2013) 2. Average employment: Annual Survey of Industries report (Government of India, 2015).

shown in Eq. (2) to be the difference between the expectation of within cluster dispersion in reference distribution, i.e., $E_n^*(\log W_k)$ and within cluster dispersion in observed data, i.e., $log\ W_k$. The null reference distribution is generated by sampling uniformly over the original dataset, out of which B Monte Carlo samples (i.e., reference datasets) are generated with each one giving within cluster sum of squares (W_{kb}^*) for $b = 1, 2\ldots. B$ and $k = 1, 2\ldots. K$. Thus, the expectation values $E_n^*(\log W_k)$ are obtained by computing the average of $log\ W_k^*$ for B Monte Carlo replicates, as shown in Eq. (3). The $log\ W_k^*$ from the Monte Carlo samples exhibit mean $(\overline{w})$ and standard deviation $sd(k)$, which, accounting for the simulation error is turned into the quantity s_k as shown in Eq. (6). The optimal number of clusters $(\widehat{k})$ is obtained as the smallest value of k for that yields the largest gap statistic G_k. This essentially indicates that the clustering structure in observed dataset is discernible from the random uniform distribution of points. Based on this method, the number of main clusters within the observed data can be determined. The extracted main clusters are then considered as subsamples, and the clustering process is repeated for them to identify the optimal number of nested clusters. That is, the subsamples (clusters) are divided into sub-subsamples (nested clusters) for revealing the dominant activity-travel patterns in each group. This procedure is termed as nested clustering because of the fact that main clusters are divided into a number of nests here.

3.2.2. Logistic regression

In this last step, the extracted freight activity-travel patterns are modeled as unordered discrete choices using multinomial logistic regression (MNL) technique. These models are founded upon random utility theory which suggests that the probability (P_{ij}) that an establishment i exhibits a particular freight activity-travel pattern j equals the probability that U_{ij} is larger than utilities U_{il} of all other alternatives in establishments' choice set C_i (0, 1, 2 … J) (Hensher et al., 2005). The MNL model is structured with the prior assumption that establishments' choice is dependent on the case-specific regressors $X_i^{'}$ (establishment size characteristics, commodity characteristics and locational characteristics), rather than the characteristics of the choice alternatives. The model coefficients are estimated by the method of maximum likelihood estimation (MLE) which essentially involves computing coefficient vector β_j that maximize the log-likelihood function given in Eq. (8). It may be noted that a numerical approximation method like Newton-Raphson method is used for computing β_j values since log-likelihood is a transcendental equation (McFadden, 1978). The resultant MLE parameter estimates β_j are used to predict the log odds of an alternative with respect to a reference a category as a linear combination of independent variables $X_i^{'}$, as shown in Eq. (9). Using the β_j values, predicted probabilities of freight activity-travel patterns are computed using Eqs. (10) and (11). Subsequently, marginal effects on probability of selecting a choice alternative j at a given value of explanatory variable X_{ik} is computed using Eq. (12). This is a powerful interpretative device which provides valuable information on the change in predicted probabilities due to unitary changes in the value of explanatory variable (Wulff, 2015).

$$\text{Log}-\text{likelihood function }(ln\ L): \ln L = \sum_{i=1}^{n}\sum_{j=0}^{J-1} d_{ij}\ ln\Pr(y_i = j \mid x_i) \tag{8}$$

$$\begin{array}{c}\text{Multinomial Logit}\\ \text{Model (MNL)}: ln\left[\dfrac{\Pr(y_i = j \mid x_i)}{\Pr(y_i = 0 \mid x_i)}\right] = ln\left(\dfrac{P_{ij}}{P_{i0}}\right) = \displaystyle\sum_{j=0}^{J-1}\beta_j X_i^{'}\end{array} \tag{9}$$

$$\begin{array}{l}\text{Predicted Probability of Reference Alternative }(P_{i0})\\ : \mathrm{P}(y_i = 0 \mid x_i) = P_{i0} = \dfrac{1}{1 + \sum_{j=1}^{J-1} exp(\beta_j X_i^{'})}\end{array} \tag{10}$$

$$\begin{array}{l}\text{Predicted Probability of all Other Alternatives }(P_{ij})\\ : \Pr(y_i = j \mid x_i) = P_{ij} = \dfrac{exp(\beta_j X_i^{'})}{1 + \sum_{j=1}^{J-1} exp(\beta_j X_i^{'})} \forall j = 1\ldots J-1\end{array} \tag{11}$$

$$\begin{array}{l}\text{Marginal Effects }(ME_{ij}): ME_{ij} = \dfrac{\partial P_{ij}}{\partial X_{ik}} = P_{ij}\left[\beta_{jk} - \displaystyle\sum_{l=1}^{J-1}\beta_{lk} P_i(l)\right] \forall j\\ \qquad = 1\ldots J\ ; \forall l(l \in J \wedge l \neq j)\end{array} \tag{12}$$

Where, $X_i^{'}$ is establishment-specific regressors thought to explain freight activity-travel patterns; β_j is the coefficient vector that contains intercept β_{0j} and coefficients β_{kj}; d_{ij} is a set of dummy variables where $d_{ij} = 1$ if $y_i = 1$ and 0 otherwise; l indicate all alternatives in J, except j.

4. Quantifying representative freight activity-travel patterns (RFAPs): case study

In this study, representative freight activity-travel patterns (RFAPs) are defined as the 'revealed pattern of freight orders by establishments over a specified time period' and used the basic unit of analysis towards developing activity-based freight transport models. Classification is involved in the categorization of establishments' activity-travel patterns into a limited number of RFAPs. The variables used to cluster establishments are: (i) annual freight production (tons); (ii) average shipment size (tons); (iii) shipment frequency (number of shipments in a week) and (iv) length of haul (average distance in kilometer travelled per shipment). As explained in the methodological framework, K-means clustering algorithm is applied to identify RFAPs within the survey sample. By developing a typology of RFAPs, this study attempts to quantify the complete daily activity-travel patterns of establishments in a holistic manner. The statistical analyses performed in this paper use R statistical computing project version 3.4.4 (R Core Team, 2017) by RStudio (RStudio Team, 2015).

4.1. Determining optimal number of clusters

The Gap statistics values of different cluster solutions are computed using '*NbClust*' (Charrad et al., 2014) and '*factoextra*' (Kassambara and Mundt, 2017) packages in R. The gap statistics (G_k) values and within cluster sum of squares (W_k) values are presented in Fig. 2 (A) as a function of the corresponding number of clusters (k). The optimal number of clusters $(\widehat{k})$ is the smallest k such that $G_k \geq G_{k+1} - s_{k+1}$, i.e., the first local maxima of G_k function in the graph. Fig. 2(A) reveals that three subgroups $(\widehat{k} = 3)$ most parsimoniously capture the underlying data structure. However, the gap function is observed to rise again after four clusters $(k = 4)$, suggesting that there are three well separated clusters and more less separated ones in the dataset. The non-monotonic behavior of gap function clearly indicates the possibility of nested clusters within the larger well separated main clusters (i.e., $\widehat{k} = 3$). Therefore, the three main clusters are considered as subsamples and the clustering process is repeated until the optimal number of nested clusters within each main cluster is identified. As shown in Fig. 2 (B), the optimal number of nested clusters within main cluster 1 is five $(\widehat{k}_{N_1} = 5)$. Similarly, Fig. 2 (D) indicate that there are three nested clusters $(\widehat{k}_{N_3} = 3)$ within main cluster 3. Instead of increasing with additional number clusters, the gap function is found to decrease from $k = 1$ to $k = 2$ in Fig. 2 (C). This suggested that there are no nested clusters $(\widehat{k}_{N_2} = 1)$ in the case of Main cluster 2. It may be noted that the total number of nested clusters $(\widehat{k}_{N_1} + \widehat{k}_{N_2} + \widehat{k}_{N_3} = 9)$ concurs with the second local maxima $(\widehat{k} = 9)$ in Fig. 2 (A) which underlines the logical validity of nested clustering procedure.

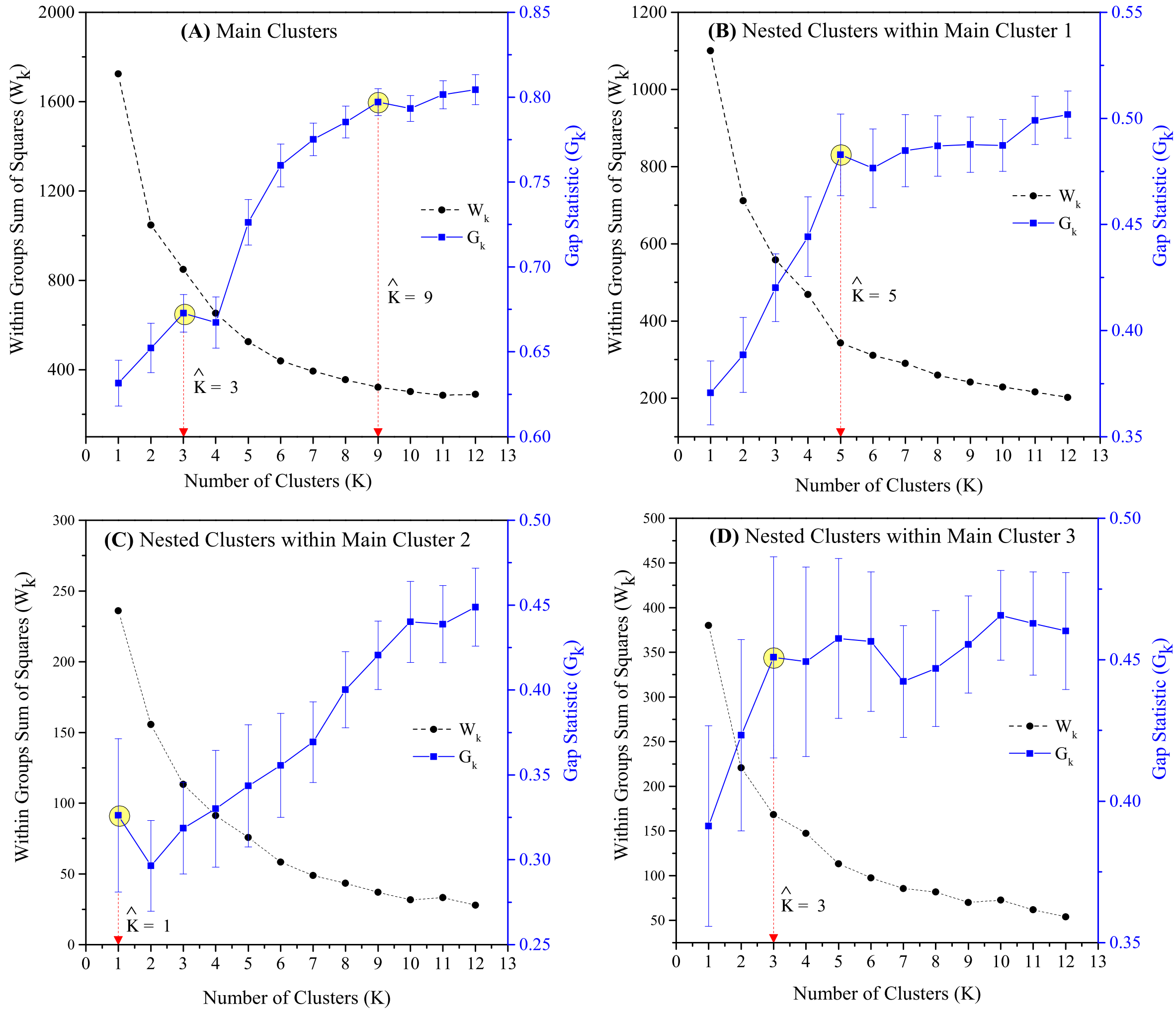

Fig. 2. Gap Statistic (G_k) within cluster sum of squares of errors (W_k) plot to determine the optimum number of clusters.

4.2. *Interpreting clusters membership characteristics and identifiable traits*

The extracted clusters of freight activity-travel patterns in main clusters and nested clusters are recognized, labeled and interpreted on the basis of their cluster centers. The proportion of establishments and the cluster centers in main clusters are presented in Table 2. Based on these values, clusters are labeled and interpretations are provided for cluster membership. The main clusters are labeled on the basis of variation in freight generation (FP) which has maximum amount of variability across cluster centers. For example, main cluster 1 is labeled as small-scale freight producers since the average freight production of establishments in this cluster is 503 tons/year. In line with this, main cluster 2 is labeled as medium-scale freight producers (FP = 1630 tons/year) and main cluster 3 is labeled as large-scale freight producers (FP = 2477 tons/year). The centers of nested clusters within main cluster 1 and 3 is given in Table 3. As shown, the nested clusters are labeled by identifying interpretable levels of remaining three variables, such as shipment size, shipment frequency and length of haul. For instance, the first variable - shipment size - is interpreted on the basis of gross vehicle weight (GVW) which determine the permitted freight weight (PFW) or “payload capacity” of trucks. Based on the GVW, trucks are classified into three in India (Pani et al., 2022): (i) light-duty vehicles (LDVs) – GVW <3.5 tons; (ii) medium-duty vehicles (MDVs) – 3.5 tons < GVW <12 tons; (iii) heavy-duty vehicles (HDVs) – GVW >12 tons. The second variable - shipment frequency - is interpreted by dividing it into three levels: (a) Low frequency (LF) - 1 to 2 shipments/week; (b) medium frequency (MF) - 3 to 4 shipments per week and (c) high frequency (HF) - 5 to 6 shipments per week. The third variable - length of haul - is interpreted by dividing it into three levels on the basis of average trucking distance covered per day in India (300 km/day) (Sople, 2010). Based on this, the length of haul is classified into: (a) Short-haul (SH) - up to 150 km; (b) Medium-haul (MH) - 150 to 300 km and (c) Long-haul

Table 2
Main cluster centers.

Clustering Variables	Group 1	Group 2	Group 3
	(63.9%)	(13.9%)	(22.2%)
	Small-Scale Freight Producers	Medium-Scale Freight Producers	Large-Scale Freight Producers
	[SSFP]	[MSFP]	[LSFP]
Freight Generation (Tons)	**503**	**1630**	**2477**
Shipment Size (tons)	4	6	19
Shipment Frequency	3	8	2
Length of Haul (Km)	146.3	240.6	245.8

Note: The figures in parentheses denote the proportion of establishments in each cluster.

Table 3
Nested cluster centers.

Clustering Variables	Group 1 *(63.9%)*					Group 3 *(22.2%)*		
	1.1	1.2	1.3	1.4	1.5	3.1	3.2	3.3
	(15.3%)	(18.1%)	(13.0%)	(9.0%)	(8.6%)	(11.3%)	(7.4%)	(3.5%)
	LDV Users with High Frequency Short Haul Shipments	LDV Users with Low Frequency Short Haul Shipments	LDV users with Medium Frequency Medium Haul Shipments	MDV Users with Medium Frequency Medium Haul Shipments	MDV Users with Low Frequency Medium Haul Shipments	HDV Users with Low Frequency Medium Haul Shipments	HDV Users with Low Frequency Long Haul Shipments	HDV Users with Medium Frequency Long Haul Shipments
	[LDV-HFSH]	[LDV-LFSH]	[LDV-MFMH]	[MDV-MFMH]	[MDV-LFMH]	[HDV-LFMH]	[HDV-LFLH]	[HDV-MFLH]
Freight Generation (Tons)	355.9	261.8	481.6	1267.9	496.5	1900.4	2166.8	4945.3
Shipment Size (Tons)	**1**	**2**	**3**	**6**	**10**	**16**	**27**	**26**
Shipment Frequency (Trips)	**6**	**2**	**3**	**4**	**1**	**2**	**1**	**3**
Length of Haul (Km/Trips)	**117.9**	**115.9**	**180.8**	**172.6**	**181.4**	**230.7**	**304.1**	**350.8**

Note: The figures in parentheses denote the proportion of establishments in each cluster.

(LH) - greater than 300 km. In order to further examine the characteristics and special features of the main clusters and nested clusters, industry sector characteristics are found for each cluster as presented in Fig. 3. The identifiable traits of these clusters are explained, as follows, using predominant industry categories that belong to each cluster.

4.2.1. Main cluster 1: small scale freight producers (SSFP)

This cluster accounts for 63.9% of the sample. From the cluster centers, it can be observed that the average annual FP of SSFP is 503 tons with an average shipment size of 4 tons, average shipment frequency of 3 trips/week and average length of haul of 146.3 km. While there is no predominant industry type associated to this cluster, establishments in this cluster mainly belong to ISIC 10 food products (23.55%), ISIC 26–28

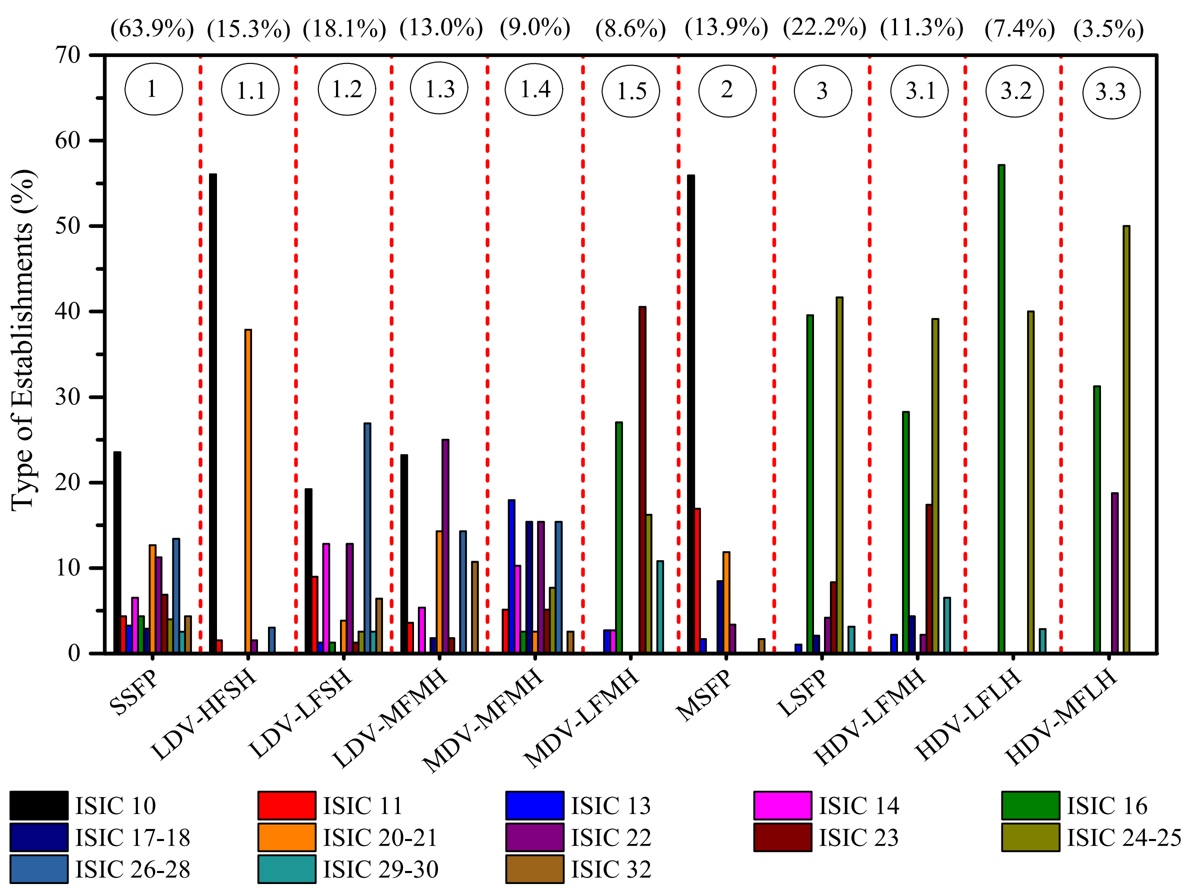

Fig. 3. Type of establishments in travel behavior clusters (The numbers in parentheses denote the proportion of establishments in each cluster; numbers in circles indicate the cluster numbers).

machinery (13.4%), ISIC 20–21 chemical products (12.68%) and ISIC 22 plastic products (11.23%).

- *Nested Cluster 1.1: LDV-HFSH* - The first nested cluster, representing 15.3% of the sample is labeled as LDV-HFSH. As suggested by the label, the establishments in this cluster are associated with an average shipment size of 1 ton (i.e., LDV), average shipment frequency of 6 shipments per week and average length of haul of 117.9 km. This activity-travel pattern with high frequency short-haul shipments suggest that products handled by establishments in this cluster are predominantly fast-moving consumer goods (FMCG) which have frequent requirements for stock-replenishment and short shelf-life. This is upheld by the fact that one of the major targeted users of vehicle type associated with this cluster, (i.e., LDVs) is transporters of FMCG products in India (Businessline, 2014). This is further indorsed by the incidence of high percentage of establishments in this cluster belong to ISIC 10 food products (56.06%), followed by ISIC 20–21 chemical products (37.88%) which handle consumable products (see Fig. 3).
- *Nested Cluster 1.2: LDV-LFSH* - This cluster accounts for 18.1% of establishments the sample. The cluster centers suggest that the establishments are associated with an average shipment size of 2 tons (LDV), average shipment frequency of 2 shipments/week (LF) and average length of haul of 115.9 km (SH). A high share of establishments in this cluster are from ISIC 26–28 machinery (26.92%), ISIC 10 food products (19.23%).
- *Nested Cluster 1.3: LDV-MFMH* - This cluster represents 13% of the sample. The cluster centers are obtained as 3 tons (LDV), 3 shipments/week (MF) and 180.8 km (MH) respectively for shipment size, shipment frequency and length of haul. The cluster is characterized by following industry sector characteristics. It has relatively large share of establishments from ISIC 22 plastic products (25%), followed by a fair share of ISIC 10 food products (23.21%), ISIC 20–21 chemical products (14.29%) and ISIC 26–28 machinery (14.29%).
- *Nested Cluster 1.4: MDV-MFMH* - This cluster, accounting for 9% of the sample, is centered with a shipment size of 6 tons (MDV), shipment frequency of 4 shipments/week (MF) and length of haul of 172.6 km (MH). While most of the ISIC groups reveal commensurate share in this cluster, establishments from ISIC 10 food products are notably absent. Apart from the ISIC groups dominant in the main cluster, establishments from ISIC 17–18 paper products (15.38%) and ISIC 14 textile products (10.26%) exhibit this activity-travel pattern.
- *Nested Cluster 1.5: MDV-LFMH* - This cluster which makes up 8.6% of the sample contains high share of establishments from ISIC 23 mineral products (40.54%) and ISIC 16 wood products (27.03%) which is not reflected in the industry sector profile of the main cluster. Besides, it is important to note that the dominant ISIC groups in the main cluster (i.e., ISIC 10, 26–28, 20–21 and 22) are absent in this nested cluster. This suggest that activity-travel patterns of establishments handling FMCG products (LDV-HFSH) is markedly distinct from nested clusters MDV-MFMH and MDV-LFMH.

4.2.2. Main cluster 2: medium-scale freight producers (MSFP) or MDV-HFMH

This cluster is the smallest (13.9% of sample) among the main clusters; it is centered at annual FP of 1630 tons/year, shipment size of 6 tons (MDV), shipment frequency of 8 trips/week (HF) and length of haul of 240.6 km (MH). This cluster may also be labeled as MDV-HFMH since there are no nested clusters within this cluster. As in the case of SSFP, ISIC 10 have the highest share (55.93%) in MSFP, while there is a fair share of establishments from ISIC 11 (16.95%) and ISIC 20–21 (11.86%). In contrast to SSFP, establishments from ISIC 26–28 are absent in this cluster and ISIC 22 plastic products occupy a limited share (3.39%). The overall industry profile suggest that products handled by MSFP are largely FMCG and, therefore, similar to LDV-HFSH within SSFP. This demonstrate that SSFP handling high frequency FMCG shipments typically prefer LDVs for short-haul trips, whereas MSFP handling the same products select MDVs for medium-haul trips. These findings underline the notion that establishments realize economies of distance by using heavier vehicles for longer trips. Given that the average shipment size increases from 1 ton (LDV-HFSH) to 6 tons (MSFP) for a commensurate increase in average annual FP of 355.9 tons–1630 tons, it may be inferred that establishments attempt to achieve the economies of scale by hauling larger quantities.

4.2.3. Main cluster 3: large-scale freight producers (LSFP)

The third main cluster is labeled as LSFP and has a share of 22.2% in the sample. The cluster is centered at annual FP of 2477 tons/year, shipment size of 19 tons, shipment frequency of 2 trips/week and length of haul of 245.8 km. The industry sector profile of LSFP has large shares of ISIC 24–25 (41.67%) and ISIC 16 (39.58%). In contrast to SSFP and MSFP, establishments handling consumer goods (ISIC 10, ISIC 11, ISIC 20–21, ISIC 26–28) are notably absent in LSFP.

- *Nested Cluster 3.1: HDV-LFMH* - This cluster accounts for 11.3% of the sample. The cluster centers are obtained as 16 tons/shipment (HDV), 2 trips/week (LF) and 230.7 km (MH). The industry sector profile of this cluster largely mirrors the relative shares of ISIC 24–25 metal products (39.13%) and ISIC 16 wood products (28.26%) in the main cluster. In addition, there is a fair share of establishments from ISIC 23 mineral products (17.39%).
- *Nested Cluster 3.2: HDV-LFLH* - According to the cluster centers, HDV-LFLH (7.4% of sample) comprises establishments producing shipment size of 27 tons (HDV), shipment frequency of 1 trip/week (LF) and length of haul of 304.1 km (LH). The industry characteristic that strongly distinguish this cluster within LSFP is the very high share of ISIC 16 wood products (57.14%). The dominant ISIC group in LSFP - ISIC 24–25 metal products - occupy high share (40%) in this cluster as well. The presence of remaining ISIC groups is negligible in this cluster.
- *Nested Cluster 3.2: HDV-MFLH* - This cluster represents the smallest segment (3.5%) of the sample. The cluster is centered at shipment size of 26 tons (HDV), shipment frequency of 3 trips/week (MF) and length of haul of 350.8 km (LH). The establishments in this cluster predominantly are from ISIC 24–25 metal products (50%), followed by ISIC 16 wood products (31.25%) and ISIC 22 plastic products (18.75%).

The schematic representation of the entire typology of freight activity-travel patterns is presented in Fig. 4. A closer look at the identifiable traits of various clusters reveal that the groups defined on the basis of *a priori* classification systems (i.e., ISIC groups) are not distinctly associated with each of the freight activity-travel pattern. Moreover, similar industry sector profiles are observed to be linked with varied freight travel patterns (LDV-HFSH and MDV-HFMH). This suggests that the commonly-used *a priori* classification systems in trip-based or tour-based freight demand modelling are overly simplified representations of the complex structure of the activity-travel patterns of establishments.

5. Freight activity-travel pattern generation (FAPG) models

Once the typology is created, the objective is to allocate the establishments to RFAP type depending on their intrinsic characteristics. To this effect, activity-travel patterns are treated as an unordered-choice mechanism which is consistent with the random utility theory. The establishments are hypothesized to exhibit the RFAPs on the basis of its fundamental desire for utility maximization since each establishment derive a particular utility value from each activity-travel pattern. The probabilities of these RFAPs are modeled using MNL model. Only the variables easily projected with population forecasts are included in the model specification and they are as follows: (a) establishment size in-

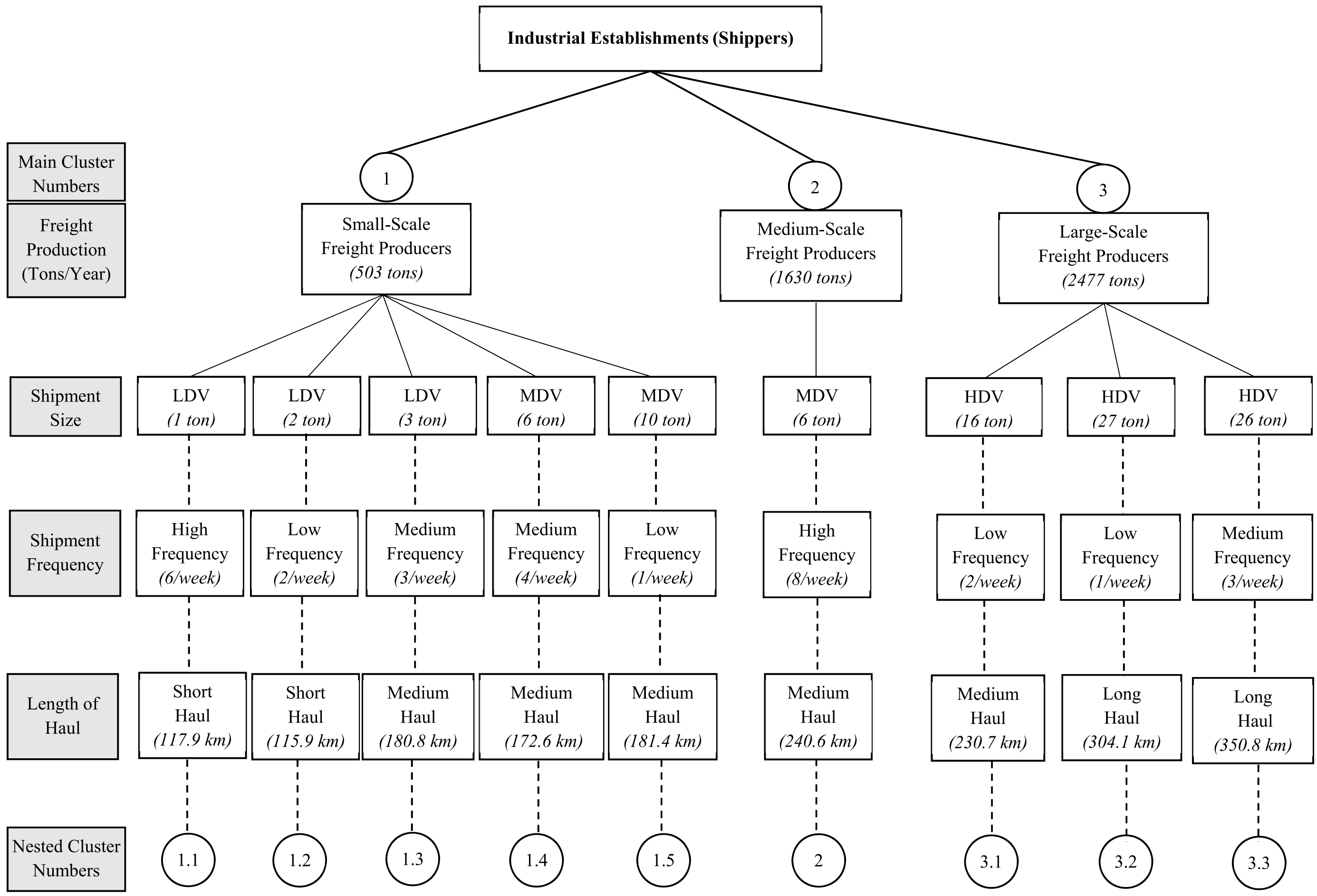

Fig. 4. Schematic representation of representative freight activity-travel patterns of establishments.

dicators (employment, gross floor area, years in business); (b) commodity characteristics (value density) and (c) locational characteristics (distance to city centroid, distance to nearest port, distance to nearest Interstate Highway). A forward sequential procedure of selection of variables is applied to build the MNL model which best defines the activity-travel patterns. That is, the variables are included in each step are selected on the basis of likelihood ratio χ^2 statistic, Akaike information criterion (*AIC*) and statistical significance of model parameters. Further, Independence from Irrelevant alternatives (IIA) assumption was tested based on Hausman-McFadden Test and found out the null hypothesis (i.e., IIA assumption is not violated) cannot be rejected since the p-values for each alternative RFAP were very high. In sum, Hausman-McFadden test suggested that IIA assumption is not violated and the estimated MNL parameters are robust. The analyses revealed that the best subset of variables is: product value density (*VD*), employment (*NE*), gross floor area (*GFA*) and distance to nearest port

Table 4
Multinomial logistic regression model parameter estimate results.

Variable (X_i')	Parameter Estimates of MNL model (β_j)							
	SSFP				MSFP	LSFP		
	LDV-LFSH	LDV-MFMH	MDV-MFMH	MDV-LFMH	MDV-HFMH	HDV-LFMH	HDV-LFLH	HDV-MFLH
Value Density (*VD*)	0.098***	0.032*	−0.068*	−0.327***	−0.266***	−0.677***	−1.505***	−0.827***
	(0.026)	(0.019)	(0.039)	(0.091)	(0.056)	(0.128)	(0.338)	(0.206)
Employment (*NE*)	−0.320**	–	0.415***	–	0.566***	0.670***	–	0.764***
	(0.155)	–	(0.143)	–	(0.143)	(0.154)	–	(0.168)
Gross Floor Area (*GFA*)	–	–	0.156***	–	0.146***	0.146**	0.205***	0.227***
	–	–	(0.058)	–	(0.056)	(0.059)	(0.066)	(0.064)
Distance to Nearest Port (*DP*)	–	–	0.052*	0.082***	–	0.072**	–	0.159***
	–	–	(0.030)	(0.030)	–	(0.031)	–	(0.048)
Intercept (β_0)	–	−1.108**	−2.983***	−1.539**	−1.630***	−2.104***	–	−5.682***
	–	(0.488)	(0.581)	(0.607)	(0.506)	(0.602)	–	(1.094)
Model Fit Statistics								
Number of observations = 432 *AIC* = 1522.219			Log-Likelihood: −721.11 Likelihood ratio (χ^2) test statistic:=393.43***				McFadden Pseudo R^2: 0.2143	

*$p < 0.1$; **$p < 0.05$; ***$p < 0.01$; #*Valued given in the parenthesis are the standard error associated with the coefficient estimates* ***Notes***: *(a) Value density is measured in 100 INR/kg; (b) Employment is measured in 10 employees; (c) Gross floor area is measured in 100 m^2; (d) Distance to nearest port is measured in* 10 km.

(*DP*). The maximum-likelihood estimates for the coefficients of these variables are calculated using '*nnet*' package (Venables and Ripley, 2003) in R. The model estimation results and goodness of fit measures are presented in Table 4.

The parameter estimates in MNL model represent the effect of explanatory variables on the utility of each activity-travel pattern alternative with respect to the reference category (LDV-HFSH). The parameter signs are consistent with the expectations. As the value density of commodity handled by the establishment increases, the establishment prefers light-duty vehicles (LDVs), as opposed to medium and heavy-duty vehicles ((MDVs and HDVs). This reflects, of course, the popular assertion that high valued products are transported in smaller quantities to save inventory holding costs (de Jong and Ben-Akiva, 2007). It may also be seen that the propensity to generate longer trips (MH or LH) reduces when the value density of the commodity increases. This may be indicative of the fact that for long distance shipments of high value, railways and inland barge shipping are competitive alternatives for road transport due to improved safety of commodities (de Jong et al., 2010). As the employment and gross floor area of an establishment increase, it is apparent that the establishments tend to exhibit travel patterns involving heavy vehicles (i.e., higher shipment size), longer trip lengths and lower frequency of shipments. These correspondences can be interpreted as follows. First of all, the pattern of higher shipment size (SS) aligns greatly with (i) theory of production functions in neoclassical economics (Besanko and Braeutigam, 2014) and (ii) economies of scale. The theory of production functions (i.e., existence of a fixed set of input variables to produce a designated quantity of outputs) holds true in the model estimation results, since larger quantity of inputs (employment, area) are indeed resulting in larger quantity of outputs (SS). The model coefficients are also in agreement with logical linkages related to economies of scale, since large establishments (MSFP and LSFP) are proportionately generating more output (i.e., SS) per unit input than small establishments (SSFP). These two reasonings may jointly explain the prevalence of higher shipment size (selection of large vehicle types) with higher employment and area levels. Higher shipment size may, in turn, lead to lower shipment frequency due to the tendency to reduce fixed costs associated with a single shipment (Békés et al., 2017). The pattern of longer trip lengths associated with these shipments, parenthetically, resonates with concept of the economies of distance (i.e., reduction in logistics cost per unit weight when the distance increases). While comparing the magnitude of coefficients, it can also be seen that the effect of employment is greater than that of gross floor area, except for HDV-LFLH alternative where the effect is statistically significant only in terms of the latter. This is synchronous with the findings of Holguín-Veras et al. (2012) which suggested that the amount of land available to an establishment acts as a constraint, rather than as an input for economic processes, thus limiting the ability of gross floor area variable to explain freight generated by establishments. The coefficient of distance to nearest port *(DP)* variable suggest that relative location of the establishment play a minor, yet, noticeable role in determining the travel patterns of establishments, although the effect is not statistically significant across all alternatives. The model clearly reveals that the establishments tend to opt for heavier vehicle types (MDVs and HDVs), longer haulages and lower frequency when the distance to the nearest port is higher. This predisposition may be attributed to the competitive advantage created for road transport over maritime transport when the ports are located far away from the establishments, especially since freight transport mode selection is strongly affected by transit time (Tiwari et al., 2003).

5.1. Predicted probabilities of freight activity-travel patterns

The preceding interpretation of coefficients in MNL model is limited to representing the contrasts among the freight activity-travel patterns, making it difficult to see the implications for each travel pattern with the changes in explanatory variables. Another issue is that unlike binary logit models, a positive sign on coefficients in an MNL model does not necessarily mean that an increase in the explanatory variable corresponds to an increase in the probability of exhibiting a particular choice alternative all the times (Wulff, 2015). Therefore, predicted probabilities $P_i(j)$ of RFAPs are computed and plotted in Fig. 5 to present an intuitive representation of the relationships between a selected explanatory variable and the predicted probabilities. A closer look at the predicted probabilities reveal that the freight travel patterns probabilities are non-linear and even non-monotonic (i.e., increasing and decreasing on different intervals of X_{ik}) in many cases. This is clearly reflected in the case of employment-based $P_i(j)$ plot (Fig. 5-B) where MDV-HFMH is the dominant alternative for establishments having 27 to 110 employees, after which the predicted probability starts decreasing. The inferences made from the MNL model coefficients are further evident while comparing the predicted probabilities of alternative RFAPs. For instance, as the value density of the commodity increases beyond its mean (545.5 INR/Kg), establishments tend to prefer LDVs over the heavier vehicles.

5.2. Marginal effects on probabilities of freight activity-travel patterns

Marginal effects on predicted probabilities are computed at representative values of explanatory variables (minimum, first quartile (Q_1), mean, second quartile (Q_2) and maximum). This is particularly relevant since predicted probabilities are non-monotonic in nature. The marginal effects presented in Fig. 6 provide quantitative assessment of the magnitude and the significance of relationship between a particular explanatory variable and choice outcomes. To illustrate this point, consider the marginal effects in terms of value density. At first quartile (Q_1), a one-unit increase in value density (corresponding to 100 INR/kg) is associated with increased likelihood of travel patterns involving light and medium duty vehicles, such as LDV-HFSH (3.75%), LDV-LFSH (1.82%), LDV-MFMH (3.2%), MDV-MFMH (3.06%), MDV-LFMH (2.69%) and MDV-HFMH (5.54%). In contrast, the likelihood of exhibiting travel patterns involving heavy duty vehicles decrease rapidly with every one-unit increase in value density, such as HDV-LFMH (−3.41%), HDV-LFLH (−15.44%) and HDV-MFLH (−1.24%). This pattern is not consistent with higher levels of value density for travel patterns types involving MDVs. In those cases, the marginal effects decrease gradually after Q_1, reverses the direction by mean of value density. Note that this interpretation is consistent with the information gained from interpretation of MNL model coefficients. In terms of employment and area, the propensity to shift to heavier vehicles is evident in the steady increase of marginal effect on probability of exhibiting MDV-HFMH and HDV-LFMH. Consistent with the prior observations regarding area being a weak descriptor of changes in freight-activity travel patterns, the employment-based marginal effects (($\pm$ 9%) on predicted probabilities are significantly higher than area-based marginal effects ($\pm$ 1%). Also, every 1 unit increase in distance to nearest port (corresponding to 10 km) is found to effectuate a steadily increasing marginal effect (0.1%–0.6%) on travel patterns involving heavier vehicles (MDVs and HDVs).

6. Research contributions and policy implications

FAPG model proposed in this study is a novel allocation model which assigns establishments to a cluster or a typology defined in terms of key freight demand parameters, so that in the future, freight demand could be predicted based on the membership to a particular cluster. This modelling approach is important for various reasons. First and foremost, this study demonstrates the development of FAPG models as a promising alternative for freight travel analysis beyond the scope of conventional trip-based freight demand models. These models make use of EBFS data and provides a powerful forecasting tool for disentangling a system as complex as freight system. This is achieved by identifying homogeneous segments of freight travel market i.e., RFAPs, by clustering EBFS data using various attributes of freight orders. At the minimum, it is possible

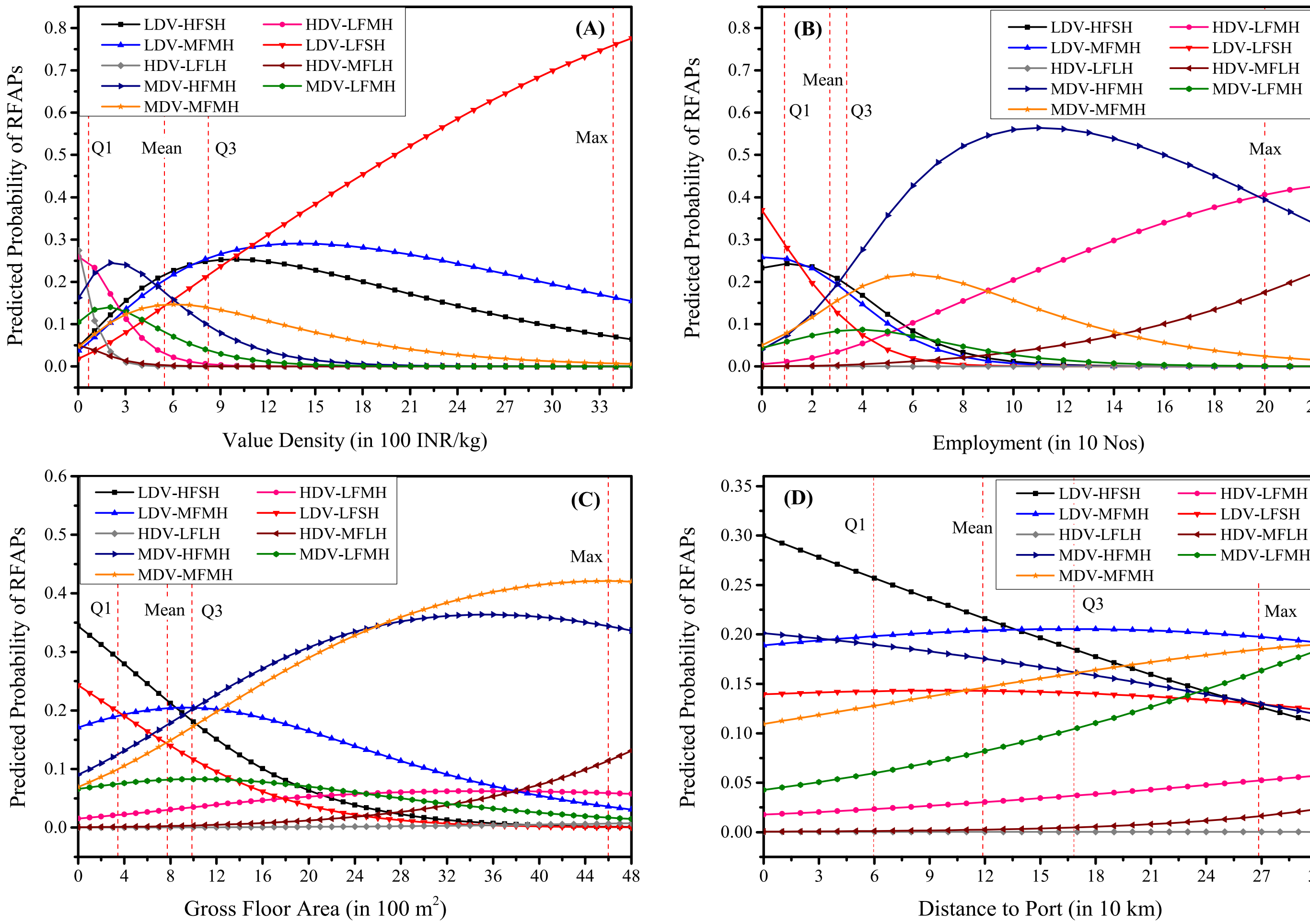

Fig. 5. Predicted probabilities of RFAPs.

that FAPG models developed using these RFAPs could replace the traditional freight generation (FG) and freight trip generation (FTG) models due to its ability to convert the assigned activity-patterns to trips or tonnage. For example, FAPG model suggest that (see Fig. 5-B) at an employment level of 120, there is a 56% probability that establishments will exhibit MDV-HFMH pattern which, in turn, implies that FP = 1630 tons/year; shipment frequency, i.e., FTP = 8 trips/week; length of haul = 240.6 km and commercial vehicle type choice = MDV. Thus, FAPG models can present an enhanced representation of freight flows since both FG and FTG are jointly modeled in this approach. That is, the best features of both commodity-based modelling (i.e., ability to capture the fundamental mechanism that drives freight demand) and vehicle-based modelling (i.e., ability to capture freight traffic implications) are included in FAPG models. Another apparent advantage is that this approach embeds the choice of commercial vehicle type in RFAPs so that FAPG models could answer the question of how commodity flows are assigned to vehicles, instead of relying upon crude approximations of payload factors. Similarly, the geographical extent of freight movements (length of haul) could also be derived from these FAPG model. Overall, this paper should prove useful in at least three fronts. The first is to forecast the differential impacts of novel freight policies and restrictions on different freight activity-travel patterns. Second, the study findings are expected to assist in identifying the variations in establishments' preferences so that it is possible to identify the type of transport supply improvements that the establishments will respond to accurately, and thus prioritize the infrastructure investments. The third area of potential application is in assisting the logistics firms for fleet size allocation and freight consolidation centers for parking space and storage unit allocation.

7. Conclusions

The conventional trip-based approach to modelling freight demand has long been criticized for its inability to reflect the underlying logistical decisions and, therefore, its inability to be responsive to demand management measures, transportation infrastructure provisions, operational planning and policy changes regarding land-use, environment and freight consolidation centers. Thus, the right direction is to take an activity-based approach to modelling freight demand with a notion that "activities" that inspire passenger trip-making is equivalent to "freight orders" in a freight system. However, there are always trade-offs involved between behavioral realism attained by a modelling approach and the resulting level of complexity in the models. That is, based on the level of granularity selected for a study, activity-based freight models could become very complicated and make its practical use limited. Besides, the proprietary nature of freight data poses significant constraints while developing activity-based freight models which are characterized with exorbitant data requirements. In this context, freight activity-pattern generation (FAPG) models developed in this study can perhaps serve as a bridge between conventional trip-based freight models and advanced activity-based freight models. These FAPG models depict logistical decisions realistically but still is simple enough to be used to predict freight demand in a study area. The central modelling object for developing FAPG models, representative freight-activity travel patterns (RFAPs), are defined using freight generation, shipment frequency, shipment size and length of haul. By identifying these RFAPs, the complexity of freight system is significantly reduced and yet, the freight activity-travel patterns are implicitly included in the model specification for FAPG models. The cluster analysis reveals that

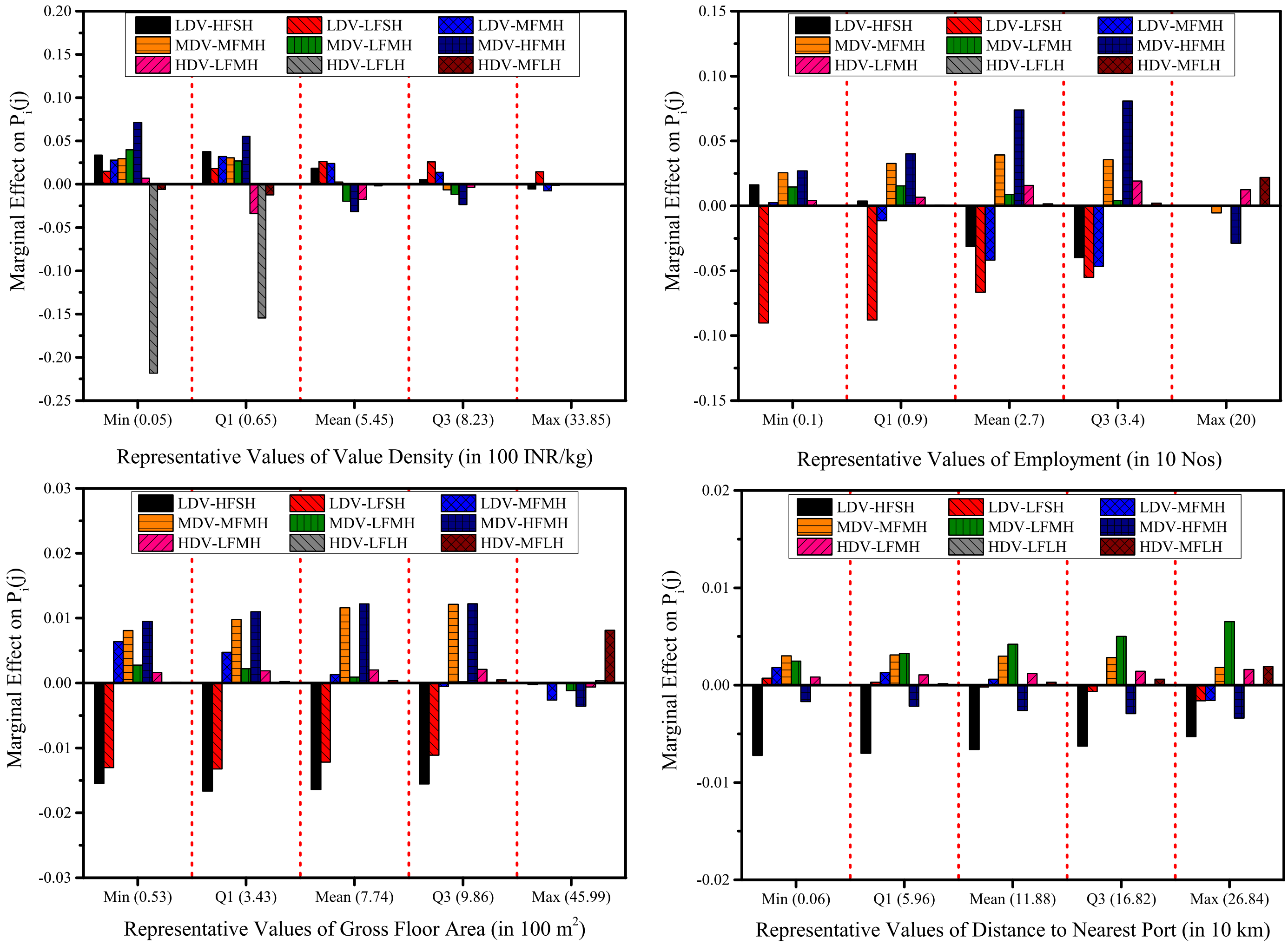

Fig. 6. Marginal effects on predicted probabilities of RFAPs.

there are three well distinguishable and nine less separated groups of RFAPs. These groups are recognizable in their cluster centers and are interpretable in terms of their industry sector profiles. The study findings do not fully substantiate the popular assertion that establishments grouped using *a priori* classification systems lead to homogeneous travel patterns. Instead, the results illustrate a rather logical contention that different establishments exhibit different travel patterns, and that, there is no one-to-one correspondence between *a priori* segments and freight activity-travel patterns. This suggests that the commonly-used *a priori* classification systems in trip-based or tour-based freight demand modelling are overly simplified representations of the complex structure of the activity-travel patterns of establishments. This finding explicitly points towards the aggregation bias that could exist in trip-based and tour-based approaches of modelling freight demand.

The main contribution of this paper is in opening the discourse for developing FAPG models based on utility maximization approach. Rather than modelling the freight 'trips' or 'tonnage' without explicit recognition of the reason for freight movement, this approach seeks to explain the probabilities of an establishment exhibiting a particular RFAP using a multinomial logit model. The results demonstrate that the freight activity-travel patterns can be predicted well in terms of establishment characteristics (employment, gross floor area), commodity value density and locational characteristics (distance to nearest port). The FAPG model coefficients, predicted probabilities of RFAPs, and marginal effects on predicted probabilities reveal interesting and interpretable findings. For example, an increase in value density is found to be associated with an increased propensity for light commercial vehicles. This pattern agrees with the inveterate truism that high valued products are transported in smaller quantities to save inventory holding costs. Similarly, as the proxy measures of business size of establishments (employment and gross floor area) increase, establishments tend to exhibit activity-travel patterns involving heavy commercial vehicles, longer haulage and lower frequency. This correspondence aligns greatly with theory of production functions and economies of distance regarding freight movements. A comparative assessment reveals that the effects of employment in determining RFAPs are typically greater than that of gross floor area. This is in line with the previous research findings that the amount of land available to an establishment may be acting as a constraint, rather than as an input for economic processes of production, thus limiting the ability of gross floor area variable to explain the activity-travel patterns of establishments. FAPG model also reveals that establishments tend to opt for heavy commercial vehicle types, longer haulages and lower frequency when the distance to the nearest port is higher, possibly due to the competitive advantage created in terms of higher accessibility and lower travel time for road transport over maritime transport in such a context.

As with any other research, this study is not without its limitations and there are several areas requiring further study. The proposed model is rather fundamental in its current form and it is hoped that the study findings can attract more research interest, especially since it only requires EBFS data, a feasible survey option for planning agencies with small budgets in the Indian subcontinent. Considering FAPG models as a base, a comprehensive activity-based freight forecasting system is worth pursuing in the future by extending the scope of EBFS with questions regarding activity-schedule, activity locations and tour characteristics. By doing so, FAPG models could have the potential to serve as the input for a comprehensive activity-based microsimulation model which can replace the conventional four step forecasting process. A future research direction would be exploring other clustering techniques such as latent class clustering or density-based clustering for defining RFAPs. The research in this direction is bound to enhance the behavioral and spatial foundations of freight models which are currently lagging in providing quantitative solutions for facilitating seamless freight movements in a highly urbanized world having a logistics-driven economy. It is also recommended to explore the transferability of this methodology to other regions and test the potential benefits in terms of model estimation and accuracy of results.

Author statement

The authors confirm their contribution to the paper as follows:
Research Conceptualization: Agnivesh Pani and Prasanta Sahu.
Data collection: Agnivesh Pani.
Analysis and interpretation of results: Agnivesh Pani, Prasanta Sahu.
Revision ideas: Agnivesh Pani, Prasanta Sahu, Lóránt Tavasszy.
Revised manuscript preparation and supplement: Agnivesh Pani, Prasanta Sahu, Lóránt Tavasszy, Sabya Mishra.
All authors reviewed the revision results and approved the final version of the revised manuscript.

Declaration of competing interest

On behalf of all authors, the corresponding author states that there is no conflict of interest.

Data availability

Data will be made available on request.

Acknowledgment

This research is funded by the Research Initiation Grant (RIG Head 06/03/302). Birla Institute of Technology and Science (BITS) Pilani, India.

References

Alho, A.R., de Abreu e Silva, J., 2016. Modeling retail establishments' freight trip generation: a comparison of methodologies to predict total weekly deliveries. Transportation 1–18. https://doi.org/10.1007/s11116-016-9700-z.

Békés, G., Fontagné, L., Muraközy, B., Vicard, V., 2017. Shipment frequency of exporters and demand uncertainty. Rev. World Econ. 153, 779–807. https://doi.org/10.1007/s10290-017-0286-0.

Besanko, D.A., Braeutigam, R.R., 2014. Microeconomics, fifth ed., fifth ed. Wiley.

Bhat, C.R., Guo, J.Y., Srinivasan, S., Sivakumar, A., 2004. Microsimulator for daily activity-travel patterns. Transp. Res. Rec. J. Transp. Res. Board 1894 57–66.

Bhat, C.R., Singh, S.K., 2000. A comprehensive daily activity-travel generation model system for workers. Transport. Res. Part A Policy Pract. 34, 1–22. https://doi.org/10.1016/S0965-8564(98)00037-8.

Bieger, T., Laesser, C., 2002. Market segmentation by motivation: the case of Switzerland. J. Trav. Res. 41, 68–76. https://doi.org/10.1177/004728750204100110.

Bowman, J.L., Ben-Akiva, M.E., 2000. Activity-based disaggregate travel demand model system with activity schedules. Transport. Res. Part A Policy Pract. 35, 1–28. https://doi.org/10.1016/S0965-8564(99)00043-9.

Buliung, R.N., Kanaroglou, P.S., 2006. A GIS toolkit for exploring geographies of household activity/travel behavior. J. Transport Geogr. 14, 35–51. https://doi.org/10.1016/j.jtrangeo.2004.10.008.

Businessline, 2014. Leyland launches LCV partner [WWW Document]. URL. https://www.thehindubusinessline.com/companies/leyland-launches-lcv-partner-price-starts-at-rs-873-lakh/article23114385.ece.

Campbell, S., Jaller, M., Sánchez-Díaz, I., Holguín-Veras, J., William, P.E., Hart, H., Lawson, C.T., 2012. Comparison between industrial classification systems in freight trip generation (FTG) modeling. In: Transportation Research Board (TRB) 91st Annual Meeting. Washington DC, United States.

Castiglione, J., Bradley, M., Gliebe, J., 2014. Activity-Based Travel Demand Models: A Primer. https://doi.org/10.17226/22357.

Central Statistics Office, 2013. Sixth Economic Census of India. New Delhi. DDI-IND-MOSPI-6-EC-2013-14.

Chandra, A., Pani, A., Sahu, P.K., Majumdar, B.B., Sharma, S., 2021a. Identifying large freight traffic generators and investigating the impacts on travel pattern: a decision tree approach for last-mile delivery management. Res. Transp. Bus. Manag. 100695 https://doi.org/10.1016/j.rtbm.2021.100695.

Chandra, A., Sharath, M.N., Pani, A., Sahu, P.K., 2021b. A multi-objective genetic algorithm approach to design optimal zoning systems for freight transportation planning. J. Transport Geogr. 92, 103037 https://doi.org/10.1016/j.jtrangeo.2021.103037.

Charrad, M., Ghazzali, N., Boiteau, V., Niknafs, A., 2014. NbClust: an R package for determining the relevant number of clusters in a data set. J. Stat. Software 61, 1–36. https://doi.org/10.18637/jss.v061.i06.

Cheng, L., Chen, X., Yang, S., Cao, Z., De Vos, J., Witlox, F., 2019. Active travel for active ageing in China: the role of built environment. J. Transport Geogr. 76, 142–152. https://doi.org/10.1016/j.jtrangeo.2019.03.010.

Chow, J.Y.J., Yang, C.H., Regan, A.C., 2010. State-of-the art of freight forecast modeling: lessons learned and the road ahead. Transportation 37, 1011–1030. https://doi.org/10.1007/s11116-010-9281-1.

Dashtestaninejad, H., Vaziri, M., Timmermans, H., 2014. Classification of rural residents on the basis of activity travel characteristics. Transp. Res. Rec. J. Transp. Res. Board 2451, 20–27. https://doi.org/10.3141/2451-03.

de Jong, G., Ben-Akiva, M., 2007. A micro-simulation model of shipment size and transport chain choice. Transp. Res. Part B Methodol. 41, 950–965. https://doi.org/10.1016/j.trb.2007.05.002.

de Jong, G., Leeds, I.T.S., Stockholm, C.T.S., Schroten, A., Delft, C.E., Essen, H.V.A.N., Delft, C.E., Otten, M., Delft, C.E., Bucci, P., 2010. The price sensitivity of road freight transport – a review of elasticities. Appl. Transp. Econ. a Manag. policy Perspect. 1–21.

Dharmowijoyo, D.B.E., Susilo, Y.O., Karlström, A., 2017. Analysing the complexity of day-to-day individual activity-travel patterns using a multidimensional sequence alignment model: a case study in the Bandung Metropolitan Area, Indonesia. J. Transport Geogr. 64, 1–12. https://doi.org/10.1016/j.jtrangeo.2017.08.001.

Diana, M., Mokhtarian, P.L., 2009. Grouping travelers on the basis of their different car and transit levels of use. Transportation 36, 455–467. https://doi.org/10.1007/s11116-009-9207-y.

Everitt, B., Landau, S., Leese, M., Stahl, D., 2011. Cluster analysis. In: Quality and Quantity, fifth ed. Wiley Series in Probability and Statistics. https://doi.org/10.1007/BF00154794.

Gonzalez-Feliu, J., Sánchez-Díaz, I., 2019. The influence of aggregation level and category construction on estimation quality for freight trip generation models. Transport. Res. Part E Logist. Transp. Rev. 121, 134–148. https://doi.org/10.1016/j.tre.2018.07.007.

Government of India, 2015. Brief Industrial Profile of Cochin. Cochin.

Gunay, G., Ergun, G., Gokasar, I., 2016. Conditional freight trip generation modelling. J. Transport Geogr. 54, 102–111. https://doi.org/10.1016/j.jtrangeo.2016.05.013.

Hanson, S., Huff, J., 1986. Classification issues in the analysis of complex travel behavior. Transportation 13, 271–293. https://doi.org/10.1007/BF00148620.

Hensher, D., Figliozzi, M.A., 2007. Behavioural insights into the modelling of freight transportation and distribution systems. Transp. Res. Part B Methodol. 41, 921–923. https://doi.org/10.1016/j.trb.2007.04.002.

Hensher, D.A., Rose, J.M., Greene, W.H., 2005. Applied Choice Analysis: A Primer. Cambridge University Press, New York.

Holguín-Veras, J., Jaller, M., Destro, L., Ban, X., Lawson, C., Levinson, H., 2011. Freight generation, freight trip generation, and perils of using constant trip rates. Transp. Res. Rec. J. Transp. Res. Board 2224, 68–81. https://doi.org/10.3141/2224-09.

Holguín-Veras, J., Jaller, M., Sanchez-Diaz, I., Wojtowicz, J., Campbell, S., Levinson, H., Lawson, C., Powers, E.L., Tavasszy, L., 2012. NCFRP 19: Freight Trip Generation and Land Use: Final Report. Washington DC, United States.

Holguín-Veras, J., Lawson, C., Wang, C., Jaller, M., González-Calderón, C., Campbell, S., Kalahashti, L., Wojtowicz, J., Ramirez, D., 2016a. NCFRP 37: Using Commodity Flow Survey Microdata and Other Establishment Data to Estimate the Generation of Freight, Freight Trips, and Service Trips. https://doi.org/10.17226/24602.

Holguín-Veras, J., Xu, N., Jaller, M., Mitchell, J., 2016b. A dynamic spatial price equilibrium model of integrated urban production-transportation operations considering freight delivery tours. Transport. Sci. 50, 489–519. https://doi.org/10.1287/trsc.2015.0616.

Holgun-Veras, J., 2002. Revealed preference analysis of commercial vehicle choice process. J. Transport. Eng. 128, 336–346. https://doi.org/10.1061/(ASCE)0733-947X(2002)128:4(336).

Hunt, J.D., Stefan, K.J., 2007. Tour-based microsimulation of urban commercial movements. Transp. Res. Part B Methodol. 41, 981–1013. https://doi.org/10.1016/j.trb.2007.04.009.

Jiang, F., Johnson, P., Calzada, C., 1999. Freight demand characteristics and mode choice: an analysys of the results of modeling with disagregate revealed preference data. J. Transport. Stat. 2, 149–158.

Joh, C.-H., Arentze, T., Timmermans, H., 2001. Pattern recognition in complex activity travel patterns: comparison of euclidean distance, signal-processing theoretical, and multidimensional sequence alignment methods. Transport. Res. Rec. 1752, 16–22. https://doi.org/10.3141/1752-03.

Kassambara, A., Mundt, F., 2017. Factoextra: extract and visualize the results of multivariate data AKassambara, A., Mundt, F.: factoextra: extract and visualize the

results of multivariate data analyses. https://cran.r-project.org/package=factoextra (2017)nalyses.
Kroesen, M., 2014. Modeling the behavioral determinants of travel behavior: an application of latent transition analysis. Transport. Res. Part A Policy Pract. 65, 56–67. https://doi.org/10.1016/j.tra.2014.04.010.
Kulkarni, A., McNally, M.G., 2001. A microsimulation of daily activity patterns. In: 80th Annual Meeting of the Transportation Research Board.
Li, R., Tong, D., 2016. Constructing human activity spaces: a new approach incorporating complex urban activity-travel. J. Transport Geogr. 56, 23–35. https://doi.org/10.1016/j.jtrangeo.2016.08.013.
Liedtke, G., Friedrich, H., 2012. Generation of logistics networks in freight transportation models. Transportation 39, 1335–1351. https://doi.org/10.1007/s11116-012-9386-9.
Liedtke, G., Schepperle, H., 2004. Segmentation of the transportation market with regard to activity based freight transport modelling. Int. J. Logist. Res. Appl. 7 (3), 199–218.
McFadden, D., 1978. Modeling the choice of residential location. Transport. Res. Rec. 72–77.
McNally, M.G., 1996. An Activity-Based Microsimulation Model for Travel Demand Forecasting.
McNally, M.G., Rindt, C.R., 2007. The activity-based approach. In: Handbook of Transport Modelling, second ed. Emerald Group Publishing Limited. https://doi.org/10.1029/93GB00468.
Mitra, R., Buliung, R.N., 2012. Built environment correlates of active school transportation: neighborhood and the modifiable areal unit problem. J. Transport Geogr. 20, 51–61. https://doi.org/10.1016/j.jtrangeo.2011.07.009.
Molin, E., Mokhtarian, P., Kroesen, M., 2016. Multimodal travel groups and attitudes: a latent class cluster analysis of Dutch travelers. Transport. Res. Part A Policy Pract. 83, 14–29. https://doi.org/10.1016/j.tra.2015.11.001.
Ortúzar, J. de D., Willumsen, L.G., 2011. Modelling Transport, Modelling Transport. https://doi.org/10.1002/9781119993308.
Pani, A., Mishra, S., Sahu, P., 2022. Developing multi-vehicle freight trip generation models quantifying the relationship between logistics outsourcing and insourcing decisions. Transport. Res. Part E Logist. Transp. Rev. 159, 102632 https://doi.org/10.1016/j.tre.2022.102632.
Pani, A., Sahu, P.K., 2019a. Planning, designing and conducting establishment-based freight surveys: a synthesis of the literature, case-study examples and recommendations for best practices in future surveys. Transport Pol. 78, 58–75. https://doi.org/10.1016/j.tranpol.2019.04.006.
Pani, A., Sahu, P.K., 2019b. Comparative assessment of industrial classification systems for modeling freight production and freight trip production. Transport. Res. Rec. 2673, 210–224. https://doi.org/10.1177/0361198119834300.
Pani, A., Sahu, P.K., Chandra, A., Sarkar, A.K., 2019a. Assessing the extent of modifiable areal unit problem in modelling freight (trip) generation: relationship between zone design and model estimation results. J. Transport Geogr. 80, 102524 https://doi.org/10.1016/j.jtrangeo.2019.102524.
Pani, A., Sahu, P.K., Majumdar, B.B., 2019b. Expenditure-based segmentation of freight travel markets: identifying the determinants of freight transport expenditure for developing marketing strategies. Res. Transp. Bus. Manag. 33, 100437 https://doi.org/10.1016/j.rtbm.2020.100437.
Pani, A., Sahu, P.K., Patil, G.R., Sarkar, A.K., 2018. Modelling urban freight generation: a case study of seven cities in Kerala, India. Transport Pol. 69, 49–64. https://doi.org/10.1016/j.tranpol.2018.05.013.
Pinjari, A.R., Bhat, C.R., 2011. Activity-based travel demand analysis. In: A Handbook of Transport Economics, pp. 213–248. https://doi.org/10.4337/9780857930873.00017.
R Core Team, 2017. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing.
Recker, W.W., McNally, M.G., Root, G.S., 1986a. A model of complex travel behavior: Part I - theoretical Development. Transport. Res. Part A Gen. 20, 319–330. https://doi.org/10.1016/0191-2607(86)90090-7.
Recker, W.W., McNally, M.G., Root, G.S., 1986b. A model of complex travel behavior: Part II-An operational model. Transport. Res. Part A Gen. 20, 319–330. https://doi.org/10.1016/0191-2607(86)90090-7.
RStudio Team, 2015. RStudio Integrated Development Environment for R [WWW Document]. http://www.rstudio.com.
Sahu, P.K., Pani, A., 2020. Freight generation and geographical effects: modelling freight needs of establishments in developing economies and analyzing their geographical disparities. Transportation 47, 2873–2902. https://doi.org/10.1007/s11116-019-09995-5.
Sakai, T., Bhavathrathan, B.K., Alho, A., Hyodo, T., Ben-akiva, M., 2018. Commodity flow estimation for a metropolitan scale freight modeling system : supplier selection considering distribution channel using an error component logit mixture model. Transportation. https://doi.org/10.1007/s11116-018-9932-1.
Samimi, A., Mohammadian, A., Kawamura, K., Pourabdollahi, Z., 2014. An activity-based freight mode choice microsimulation model. Transp. Lett. 6, 142–151. https://doi.org/10.1179/1942787514Y.0000000021.
Sánchez-Díaz, I., Holguín-Veras, J., Ban, X.J., 2015. A time-dependent freight tour synthesis model. Transp. Res. Part B Methodol. 78, 144–168. https://doi.org/10.1016/j.trb.2015.04.007.
Sánchez-Díaz, I., Holguín-Veras, J., Wang, X., 2016. An exploratory analysis of spatial effects on freight trip attraction. Transportation 43, 177–196. https://doi.org/10.1007/s11116-014-9570-1.
Sople, V.V., 2010. Logistics Management: the Supply Chain Imperative. Dorling Kindersley (India).
Tavasszy, L., De Jong, G., 2014. Modelling Freight Transport, First. Elsevier. Elsevier, London. https://doi.org/10.1016/B978-0-12-410400-6.00008-2.
Tibshirani, R., Walther, G., Hastie, T., 2001. Estimating the number of clusters in a data set via the gap statistic. J. R. Stat. Soc. Ser. B (Statistical Methodol. 63, 411–423. https://doi.org/10.1111/1467-9868.00293.
Tiwari, P., Itoh, H., Doi, M., 2003. Shippers' Port and carrier selection behaviour in China: a discrete choice analysis. Marit. Econ. Logist. 5, 23–39. https://doi.org/10.1057/palgrave.mel.9100062.
Venables, W.N., Ripley, B.D., 2003. Modern Applied Statistics with S, Fourth. Technometrics. Springer, New York. https://doi.org/10.1198/tech.2003.s33.
Wang, Q., Holguín-Veras, J., 2008. Investigation of attributes determining trip chaining behavior in hybrid microsimulation urban freight models. Transp. Res. Rec. J. Transp. Res. Board 2066, 1–8. https://doi.org/10.3141/2066–01.
Wen, C.H., Koppelman, F.S., 2000. A conceptual and methdological framework for the generation of activity-travel patterns. Transportation 27, 5–23. https://doi.org/10.1023/A:1005234603206.
Wulff, J.N., 2015. Interpreting results from the multinomial logit model: demonstrated by foreign market entry. Organ. Res. Methods 18, 300–325. https://doi.org/10.1177/1094428114560024.
Yasmin, F., Morency, C., Roorda, M.J., 2017. Trend analysis of activity generation attributes over time. Transportation 44, 69–89. https://doi.org/10.1007/s11116-015-9624-z.