



## A behavioral probabilistic model of carrier spatial repositioning decision-making

Assia Boumahdaf, Michel Broniatowski, Emilie Miranda, Antoine Le Squeren

### ► To cite this version:

Assia Boumahdaf, Michel Broniatowski, Emilie Miranda, Antoine Le Squeren. A behavioral probabilistic model of carrier spatial repositioning decision-making. 2022. hal-03744861

**HAL Id: hal-03744861**

**<https://hal.science/hal-03744861>**

Preprint submitted on 4 Aug 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# A behavioral probabilistic model of carrier spatial repositioning decision-making

Assia Boumahdaf<sup>\*</sup>, Michel Broniatowski<sup>1</sup>, Émilie Miranda, and Antoine Le Squeren

<sup>1</sup>*Laboratoire de Probabilité, Statistique et Modélisation (LPSM), Sorbonne Université, 4 place Jussieu 75005 Paris, France*

<sup>\*</sup>*Corresponding author: Assia Boumahdaf, assia.boumahdaf@gmail.com*

## Abstract

This paper studies the truckload market with carriers providing transport services between two locations. It aims to provide a modeling methodology to represent the spatial behavior of a carrier dealing with the issue of repositioning. Indeed, due to the imbalance of trade, carriers face the difficulty of finding freight for their return trips. When they operate over long distance shipments, repositioning their empty vehicles from the low-demand zone is necessary to sustain their business. Yet the mechanisms at stake by carriers to understand their repositioning decision-making process are mostly unknown and unobservable. This lack of data on carrier repositioning zone choice issues has major consequences for shipper and forwarder resource planning systems. Indeed, repositioning behavior induces hidden costs which makes it difficult to design for example a cost-based pricing strategy. To address this problem we develop a mathematical model to study the spatial repositioning behavior of carriers. We propose a probabilistic approach based on aggregated transport data that consists in a two-steps decision making process. The first one is the probabilistic selection of a set of repositioning candidates based on the microeconomic theory of the consumer. The second step is the choice of a region within this set through the estimation of the spatial distribution of reloading. It makes use of the graph structure of the transport data and combines a spatial interaction model and a random walk model on a graph. Using simulations, we illustrate how our methodology can be used for operational purposes to provide more transparency on carrier behavior. In conclusion, research perspectives are suggested for tackling the problem of freight demand estimation as well as rationalizing the impact of the trade imbalance on the price of a transport. Software development perspectives will also be addressed.

**keywords :** Imbalance of trade, carrier spatial behaviour, trip-chaining behavior, repositioning probability, random walk on bipartite graph, spatial modeling.

## 1 Introduction

### Context and problematic

The truckload market is a strategic sector of the economy and an essential link in the supply chain. According to Eurostat, road freight transport contributes to 10% of French GDP and generates a revenue of 45.8 billion euros. However, the development of road transport is facing a number of limitations in terms of technological development which correspond to the main challenges the sector has to face in order to pursue its transformation. Among these challenges, the place of digitalization and the visibility of transport are at the core of the major concerns. The rationalization and efficiency of the sector depend on the digital transition. This is increasingly important as supply chain management becomes more complex, shippers' requirements in terms of quality of service and tracking getting higher and have been intensified during the health crisis. This need for visibility includes an understanding of market dynamics and how these affect the spatial behavior of carriers.

In transport services, vehicles have to move back and forth between selected loading and unloading locations in different areas. Therefore, services in both directions can be seen as a joint production process. Indeed, if a transport company provides a service in one direction, it will have to provide the service in the opposite direction. For example, in the case of a round trip between two locations  $A$  and  $B$ , a transport movement from  $A$  to  $B$  inevitably creates a movement from  $B$  to  $A$  as the driver and the truck have to return to the original depot. The joint nature of transport is well known in economic literature and has been studied by many authors. It would seem that the pioneers were [Pigou et Taussig \(1913\)](#). We can also mention [Pederson \*et al.\* \(1979\)](#) and more recently [Fan \*et al.\* \(2014\)](#). Moreover, it is well known that trade is subject to imbalances: some areas are mainly exporting areas, whereas others are mainly importing areas.

The joint nature of transport together with the imbalance of trade generates a major logistical challenge: the problem of finding return shipments. This phenomenon is known in transport economics as the backhaul problem. It refers to the situation where the volume of transported goods (or people) is not balanced between two areas, which means that the transport flows are mainly in one dominant direction, the so-called headhaul, while the opposite direction is the so-called backhaul. To sustain their activity, carriers have to commit to a maximum transport capacity for a round-trip, and are therefore forced to reposition their empty vehicles looking for a return shipment to their originating depot. Repositioning is defined as the movement of empty vehicles from a delivery location to pickup location. The problem of vehicle repositioning is deemed as an inefficiency of the road transport sector, but it is an inevitable consequence of the trade imbalance and is very hard to be eliminated, and thus requires consideration in addressing transport issues.

In the traditional freight road market, different actors interact, such as shippers, carriers and forwarders. In this highly fragmented market which has not been fully enriched by new technologies, collaboration between the different actors and information sharing is difficult, resulting in a lack of transport data. In particular, shippers and forwarders have limited or no visibility on carriers' choices and decisions on repositioning locations. Understanding these choices is essential in determining a pricing strategy for a transport operation as there are hidden costs associated with vehicle repositioning.

There are several research streams on the backhaul problem: study of the impact of trade imbalance on transportation pricing, optimal repositioning strategy, collaborative repositioning strategy. However, to the best of our knowledge, none of them deals with the study of the impact of freight imbalance on the spatial repositioning behavior of carriers. There is a lack of analysis of carriers behavior faced with trade imbalance in order to understand their spatial choices of repositioning locations. One would need to know both the degree of imbalance and the preferences of each individual carrier, which are difficult to observe. Due to the lack of data, an aggregated approach based on annual observations of transport flows will be used to address the problem of spatial repositioning choice. Our research aims to provide an understanding of this issue by developing a mathematical model to study the decision-making and spatial choice process of carriers at an aggregate level. Decision-making is not straightforward because the choice problem involves knowledge of the market and its imbalances. To characterize the trade imbalance we develop a methodology to estimate the spatial probability of reloading. It will be defined as the probability of finding a return shipment to its depot from a region of low demand. This indicator will be at the core of the carrier choice process. The methodology is based on a random walk model on a bipartite graph. The transition matrix of the underlying Markov chain is estimated by calibrating an extension of a spatial interaction model based on origin-destination matrices. We hope that the knowledge obtained from the study of this problem can help practitioners to better understand the transport market and shed light on more practical problems.

## Our contributions

1. We consider a spatial choice problem from the point of view of carriers that has never been addressed in the literature. This problem could have an impact on the practical difficulties encountered by transport practitioners, particularly on the problems of information feedback and lack of visibility (especially the hidden costs due to empty repositioning). We develop a two-

stage decision-making process scheme including a method for estimating the spatial reloading probability;

2. We define the spatial distribution of the reloading probability across the territory; an original probabilistic indicator to characterize the trade imbalance. A methodology for estimating this spatial probability is proposed. It is based on the modelling of a random walk on a bipartite graph;
3. We develop an extension of Huff's spatial interaction model to model conditional destination choices. This representation is particularly suitable for modelling trip-chaining;
4. The distance in the previous Huff model will be defined by the transportation cost. We take into account uncertainties in the modelling of this cost based on partial information from surveys;
5. We perform numerical experiments to demonstrate how the method could be used in practice to overcome the problem of information visibility, especially in terms of hidden costs due to repositioning.

## Outline of the paper

The paper is organized as follows. In Section 2 we present the literature review pertaining to the empty equipment repositioning problem. Section 3 focuses on the necessary mathematical background on random walks on finite graphs, the spatial interaction Huff model and presents some concepts on origin-destination matrices. Section 4 presents the probabilistic choice repositioning problem by a description which relies on the mathematical framework of Section 3. It also presents the assumptions leading to the description of this model. Section 5 is dedicated to modelling the probabilistic choice problem of carrier's repositioning introduced in Section 4. Section 6 is devoted first to some preliminary tools for modelling random walks on a bipartite graph, then to the probabilistic spatial interaction model and the computation of the spatial reloading probability. Section 7 is devoted to application of our methodology by numerical experiments, and Section 8 is dedicated to the conclusion of future works. Technical details and further analyses are gathered in Appendix A for some basic concepts on statistical estimation based on  $\phi$ -divergences, Appendices B and C for the estimation procedure, and Appendix D for the stochastic transportation cost.

## 2 Literature review

Many studies on trade imbalance in the transportation industry focus on the empty equipment repositioning (EER) problem due to its importance to the transport operations management. Indeed, imbalanced trade has resulted in considerable costs and decreased profits for transportation firms. Many authors have investigated strategies to reduce the cost of repositioning empty containers. Song et Carter (2009) investigated the impact of route-coordination and container-sharing on the empty container movements. Many authors Özlem Ergun *et al.* (2007); Özener et Ergun (2008); Bailey *et al.* (2011); Pan *et al.* (2019); Ferrell *et al.* (2020) have studied how to reduce costs through collaboration network between shippers and carriers.

The EER problem is often considered as an inefficiency in the transportation industry and has attracted considerable attention and many classes of problems have been investigated. This problem exists widely in all freight transport sectors, such as trucking, rail and an extensive literature exists in the shipping industry referred to as empty container repositioning problem (ECR). Dejax et Crainic (1987) proposed an extensive review of the models dealing with empty flows and backhaul problems in the freight transportation industry.

The first stream of work focuses on the optimal repositioning strategy given the imbalanced flow of demand. In the shipping industry considerable work has been done. Crainic *et al.* (1993) proposed a dynamic deterministic formulation for the empty container allocation problem in an inland transportation system and extended it to a two-stage stochastic programming formulation under the

demand uncertainties. Song (2007) provided an optimal policy for empty container repositioning with uncertain demand using a Markov decision process. Erera *et al.* (2009) adopted robust optimization techniques to address the dynamic of the empty repositioning problem modeled using time-space networks. Long *et al.* (2012) have taken into account the uncertainty in vessels weight and space capacity and solve the problem using a two-stage stochastic program. Li *et al.* (2004) studied the empty container allocation in a port with the aim to reduce redundant empty containers. They consider the problem as a nonstandard inventory problem with simultaneous positive and negative demand under a general holding cost function. Also, Li *et al.* (2007) extended the model of Li *et al.* (2004) to a model for a multi-port case. In the railway networks we can mention Holmberg *et al.* (1998); Joborn *et al.* (2004)

Another strand of the literature has more specifically investigated the pricing decisions with empty equipment repositioning in a transportation market. Gorman (2001, 2002) studied a freight carrier's pricing strategy in a freight rail network. Equipment repositioning is required if the demand flow in the network is unbalanced. The objective is to find a set of prices that maximize the network profits. The author formulated a carrier's pricing problem as a mathematical programming model in a network and provided an efficient computational algorithm to solve the problem. Zhou et Lee (2009) considered two firms competing with each other to increase their profits from transportation services they provide between two port services. In this study, the realized demand for the services is deterministic. The authors found the optimal prices with EER of the two firms in each direction using a monopoly and a Bertrand duopoly model. Chen *et al.* (2016) extended the study of Zhou et Lee (2009) and discussed the carrier pricing decisions with EER who provide transportation services for product and waste between two port services. Xu *et al.* (2015) studied a sea-cargo service chain with one carrier and two forwarders providing transportation service between two ports with deterministic demands in both directions. They built a mathematical model to study how the carrier and forwarders determine pricing and EER cost-sharing decisions. Lu *et al.* (2020) investigated simultaneous pricing and empty container repositioning decisions considering stochastic demand in two-depot shipping services. They solved the problem by developing a large-scale dynamic programming model.

Although there are numerous studies dealing with empty equipment repositioning, to our knowledge, our work is the first one to study the impact of the trade imbalance on the spatial behavior of carriers faced with repositioning in the truckload industry. Our research therefore aims to provide an understanding of this issue by constructing and analyzing a mathematical model. However, another stream of research is related to our study, and concerns in a general way the modeling of the geographical choices of the consumer without however considering the repositioning behavior of a carrier. This stream of research is widely used in the Anglo-Saxon geography community (Thill et Timmermans (1992)). The modeling of geographic choices aims to analyze spatial decisions and consumer choice processes. The behavioral approach of the consumer in his purchasing behavior can be modeled by a trip-chaining. This type of model was conceived in the 1960s. Thill et Thomas (1987) proposes a synthesis of the existing literature on multi-stop travel behavior. This approach lends itself naturally to Markovian analysis. It allows for example to model the chain behavior of the stops during a trip (Lerman (1979)). The Markovian approach has the advantage of considering the interactions between the different stops, but it is not a choice model since it does not incorporate a preference structure. Thus, several authors have tried to take into account the choice processes of a consumer. Ben-Akiva *et al.* (1978) uses a Multinomial Logit model to estimate the transition matrix. Borgers et Timmermans (1986) use a gravity model to estimate the transition matrix of consecutive pedestrian stops in their shopping behavior. Our research therefore aims to provide an understanding of this problem through the construction and analysis of a mathematical model.

### 3 Mathematical background

In this section we specify the mathematical background that we use in this paper. As we will see in Section 4, an appropriate description of the carrier probabilistic repositioning choice problem involves the concept of a random walk on a bipartite graph. The transition matrix of the underlying Markov

chain will be calibrated using an Huff-type spatial interaction model on the basis of origin-destination matrices. We first recall some basic notions from graph theory. We will then describe the random walks on bipartite graphs. We will also review the original formulation of the Huff model, and finally, we present some basic notions related to origin-destination matrices.

### 3.1 Some basic facts on random walks on graphs

A random walk is a random process that describes paths consisting of a succession of random steps on some mathematical space (in our case a graph). Let us start by restating some graph theory definitions. In this paper, only directed and weighted graphs will be considered.

A bipartite graph is a graph having two kinds of vertices, and whose edges are only between vertices of different kinds. More formally, a graph is bipartite if the set of vertices  $V$  can be partitioned into two sets  $A$  and  $B$  such that there is no edge between the vertices of the same set. In other words, for each edge  $(u, v)$  either  $u \in A$  and  $v \in B$ , or  $u \in B$  and  $v \in A$ . We denote  $G = (A, B, E)$  such a graph, where  $A$  and  $B$  constitute two distinct classes of vertices, with  $A \cap B = \emptyset$ , and  $E \subseteq (A \times B) \cup (B \times A)$ . Let  $\mathbf{W} := (w_{ij})_{u_i \in A, v_j \in B}$  be a graph weight matrix where the entry  $w_{ij}$  is the weight of edge  $(u_i, v_j)$  appearing between the vertex  $u_i \in A$  and the vertex  $v_j \in B$ . Then the adjacency matrix  $\mathbf{A}$  of the bipartite graph is expressed as

$$\mathbf{A} = \begin{pmatrix} \mathbf{0} & \mathbf{W} \\ \mathbf{W}^t & \mathbf{0} \end{pmatrix}, \quad (1)$$

where the matrix  $\mathbf{W}^t := (\tilde{w}_{ij})_{v_i \in B, u_j \in A}$  designates the transpose of  $\mathbf{W}$ , with  $\tilde{w}_{ij} = w_{ji}$ , and corresponds to weights of edges from nodes of type  $B$  to those of type  $A$ .

Consider a weighted bipartite graph  $G = (A, B, E)$  with the associated weighted matrix (1). Let us imagine a walker on an initial vertex of the graph, say  $u_0 \in A$ , and who moves at each time step, following a randomly chosen edge. No edge connects the vertices of the same class, so the random walker, initially in  $u_0 \in A$ , randomly chooses a neighbor of  $u_0$  in  $B$  (let's denote  $v_1$  such node) and then moves on it. The walker randomly chooses again a neighbor of  $v_1$  in  $A$ , say  $u_2$ , and moves again on it, and so on... At each step  $n$  of its path, we have a random variable  $X_n$  taking values in  $A \cup B$ . The sequence of vertices  $X_1, X_2, \dots, X_k, \dots$  selected in this way by the walker is a random walk on the graph  $G$ . The random sequence of vertices is a discrete Markov chain with state space  $A \cup B$  and matrix of transition probabilities  $\mathbf{P}$ .

The initial vertex  $u_0$  can be fixed or can be a random variable of initial law  $\mu$  on  $A \cup B$ . Starting from  $u_i \in A$ , the Markov chain is characterized by the transition probabilities

$$\mathbb{P}(X_{n+1} = v_j | X_n = u_i) = \begin{cases} w_{ij}/W(u_i) & \text{if } (u_i, v_j) \in A \times B, \\ 0 & \text{otherwise,} \end{cases} \quad (2)$$

where  $W(u_i) := \sum_{v_j \in \Gamma(u_i) \subseteq B} w_{ij}^A$  corresponds to the sum of weights in the neighborhood of the vertex  $u_i \in A$ , with  $\Gamma(u_i)$  denotes the neighborhood of the vertex  $u_i$ .

Starting from  $v_i \in B$  the transition probabilities are given by

$$\mathbb{P}(X_{n+1} = u_j | X_n = v_i) = \begin{cases} \tilde{w}_{ij}/W(v_i) & \text{if } (v_i, u_j) \in B \times A, \\ 0 & \text{otherwise,} \end{cases} \quad (3)$$

where  $W(v_i) := \sum_{u_j \in \Gamma(v_i) \subseteq A} \tilde{w}_{ij}$  is the sum of weights in the neighborhood of the vertex  $v_i \in B$ , with  $\Gamma(v_i)$  denotes the neighborhood of the vertex  $v_i$ .

The transition matrix  $\mathbf{P}$  of the Markov chain  $\{X_n, n \geq 0\}$  on  $A \cup B$  with initial law  $\mu$  is given by

$$\mathbf{P} = \begin{pmatrix} \mathbf{0} & \mathbf{P}^{(A)} \\ \mathbf{P}^{(B)} & \mathbf{0} \end{pmatrix}, \quad (4)$$

where submatrices  $\mathbf{P}^{(A)}$  and  $\mathbf{P}^{(B)}$ , are given respectively by the transition probabilities (2) and (3).



### 3.2 Consumer spatial behaviour models. The formulation of the Huff model

Spatial analysis models are used to understand the spatial choice behavior of a consumer towards a supply location. Consumer choices are generally the result of a trade-off between both attractiveness and distance of the supply location. This trade-off is represented by a utility function. The first retail trading area models due to [Reilly \*et al.\* \(1929\)](#), [Reilly \(1931\)](#), or [Converse \(1949\)](#) were deterministic, and [Huff \(1964\)](#) was the first to formulate these models on a probabilistic basis. The model is based on the premise that the probability that an individual confronted with a set of alternatives will choose a particular location is directly proportional to the utility of each alternative. The Huff model is based on the following utility function

$$U_{ij} = A_j^\alpha / d_{ij}^\beta,$$

where  $U_{ij}$  is the utility of the supply location  $j$  for a customer located in  $i$ ;  $A_j$  reflects the attractiveness of the supply location generally measured by its sales area; and  $d_{ij}$  the distance of the shop from the potential customer of origin  $i$ . The parameter  $\beta$  represents the distance friction, and  $\alpha$  measures the attractiveness elasticity. The level of consumer satisfaction is positively correlated to the attractiveness of the supply location, hence  $\alpha > 0$ . On the other hand, the more distant the supply location is, the less satisfied the potential customer is, hence  $\beta > 0$ .

From the choice axiom of [Luce \(1959\)](#), the consumer's decision in the [Huff \(1964\)](#) model is considered as probabilistic. The probability that an individual  $i$  will select the alternative  $j$ , from a set of alternatives ( $j = 1, \dots, n$ ) corresponds to the utility of the supply location  $j$  ( $U_{ij}$ ) renormalized by the sum of the utilities of all the alternatives, namely

$$p_{ij} = \frac{U_{ij}}{\sum_{k=1}^n U_{ik}} = \frac{A_j^\alpha / d_{ij}^\beta}{\sum_{k=1}^n A_k^\alpha / d_{ik}^\beta}, \quad (5)$$

where  $i$  is the origin location of the potential customer, and  $j$  the supply location;  $p_{ij}$  is the probability that the consumer originated from  $i$  selects the alternative  $j$ , with  $\sum_{j=1}^n p_{ij} = 1$ , and  $0 < p_{ij} < 1$ . Both  $\alpha$  and  $\beta$  are unknown parameters which are usually estimated by linear multiple regression. In [Section 6.2](#) we propose an adaptation of the Huff Model(5) to characterize the conditional choice of destinations; decision-making that focuses on loading or unloading locations.

### 3.3 Some definitions in the origin-destination matrices framework

In this section we give a brief description of the origin-destination (OD) matrices data used in this paper. We then present some basic results on OD matrices.

#### 3.3.1 Available data

The main source of data (French and European) of freight road transport is the SitraM database (Système d'Information sur les TRAnsports de Marchandises). It was built from surveys whose observation units are based on vehicle movements of carriers. These surveys consist of a sample of road vehicles (trucks, trailers and semi-trailers) less than 15 years old and with at least 3 tonnes of capacity.

The dataset we study comes from the European freight road transport surveys conducted between 2010 and 2016. They consist of annual traffic matrices, providing an estimation of the transported goods measured in tonnes from one division to another one based on surveys of transporters. The variables used are:

1. the origins (NUTS);
2. the destinations (NUTS);
3. transported goods (ton).

### 3.3.2 Some basic facts on OD matrices

The general structure of the origin-destination (OD) matrix consists of an  $m \times n$  matrix. The intersection of the  $i$ -th row (origin) and the  $j$ -th column (destination) corresponds to a cell of the matrix with a value  $T_{ij}$ . This value provides information on the road transported goods from the origin  $i$  to the destination  $j$ . This matrix represents freight demand generally measured in tons per kilometer or tons among European NUTS regions (Nomenclature of Territorial Units for Statistics).

We restrict our study to the case of interregional trades measured in tons, and where origins and destinations are identical. This yields the square matrix that will be noted  $\mathbf{T}$  from now on, with  $\mathbf{T} := (T_{ij})_{1 \leq i, j \leq L}$ , where  $L$  denotes the number of origin (or destination) areas, i.e.,

$$\mathbf{T} = \begin{pmatrix} 0 & T_{12} & T_{13} & \cdots & T_{1L} \\ T_{21} & 0 & T_{23} & \cdots & T_{2L} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ T_{L1} & T_{L2} & T_{L3} & \cdots & 0 \end{pmatrix}, \quad (6)$$

with a zero diagonal for within-region trades.

From (6) we can derive a variety of indicators. We will mention the ones we will use in this paper. Let us begin by the total value of transported goods, i.e.,

$$T_{\bullet\bullet} = \sum_{i=1}^L \sum_{j=1}^L T_{ij}. \quad (7)$$

The total value of the emitted goods from  $i$ , denoted by  $T_{i\bullet}$ , and the total value of received goods by  $j$ , denoted by  $T_{\bullet j}$ , are respectively given by

$$\text{For all } i = 1, \dots, L, T_{i\bullet} = \sum_{j=1}^L T_{ij} \text{ and, for all } j = 1, \dots, L, T_{\bullet j} = \sum_{i=1}^L T_{ij}. \quad (8)$$

From (7) and (8), we derive the following frequency indicators

1. the proportion of total emitted goods from  $i$  relative to overall transported goods

$$E_i = T_{i\bullet}/T_{\bullet\bullet}, \quad i = 1, \dots, L \quad (9)$$

2. the proportion of total received goods in  $j$  relative to the overall transported goods

$$R_j = T_{\bullet j}/T_{\bullet\bullet}, \quad i = 1, \dots, L. \quad (10)$$

From Eqs. (9) and (10), we derive the following conditional frequency indicators. For all  $i = 1, \dots, L$ ,

$$F_{j|i} = T_{ij}/T_{i\bullet}, \quad j = 1, \dots, L; \quad (11)$$

and, for all  $j = 1, \dots, L$ ,

$$F_{i|j} = T_{ij}/T_{\bullet j}, \quad i = 1, \dots, L. \quad (12)$$

**Remark.** An OD-matrix can be considered as a contingency table. Indeed, each cell of (6) represents the value of the relationship between two qualitative characters  $X$  and  $Y$ . The two qualitative characters correspond to the origin and destination zones respectively. The intersection of the  $i$ -th modality of  $X$  with the  $j$ -th modality of  $Y$ ,  $T_{ij}$ , corresponds to the transported goods from the origin  $i$  to the destination  $j$ . From (6) and (7) we can generate a new matrix of frequencies  $\mathbf{F} := (F_{ij})_{1 \leq i, j \leq L}$ , such that,  $F_{ij} := T_{ij}/T_{\bullet\bullet}$  for  $i, j = 1, \dots, L$ . Thus,  $F_{ij}$  can be interpreted as the probability of loading goods from  $i$  and unloading goods in  $j$ , i.e.,  $(F_{ij})_{i \in \mathcal{L}, j \in \mathcal{L}}$  can be understood as a probability distribution on the probability space,  $\mathcal{L} \times \mathcal{L}$ , with  $\mathcal{L} = \{1, 2, \dots, L\}$  designates the space of origins (and destinations). By analogy with (8),  $(F_{i\bullet})_{i \in \mathcal{L}}$  and  $(F_{\bullet j})_{j \in \mathcal{L}}$  designate the marginal distributions on  $\mathcal{L}$ , and Eqs. (11)-(12) conditional distributions. Eq. (11) may be interpreted as the conditional probability of unloading goods in  $j$  given that goods were loaded from  $i$ , and Eq. (12) the conditional probability of loading from  $i$  given that the unloading area is  $j$ . The two last Equations are commonly called line profiles and column profiles.



## 4 Problem definition and methodology

This section focuses on introducing the problem definition and the methodology. We first outline the assumptions used for the modelling of the probabilistic repositioning choice problem.

### 4.1 Model assumptions

**Assumption 1** (Geographical structure). *The geographical range of the model is on within-national trades. The territory in which carriers make shipments is divided into geographical zones defined by NUTS regions. Each zone is characterized by its centroid.*

**Assumption 2** (Transportation market). *We only consider the long-haul and full truck load (FTL) transportation service since the issue of repositioning is mostly relevant for long-haul shipments. The definition of long distance used is the one defined by the French National Road Comity (CNR), i.e., a transport where operating constraints make impossible or uncertain for the driver to return home every day. The truck capacity is assumed to be either 0 (empty truck) or  $Q$  (fully loaded truck). A vehicle will be considered as a fully loaded truck if its capacity  $Q$  is such that  $Q \in [Q_{\min}, Q_{\max}]$ .  $Q_{\min}$  is the minimum limit for considering a truck to be fully loaded.  $Q_{\max}$  is the maximum allowed weight for a truck.*

**Assumption 3** (Long-haul repositioning). *Repositioning behavior is defined as the movement of an empty vehicle from an unloading zone to a (re)loading zone. In the long-haul market, we assume that the (re)loading zone is different from both the unloading and loading zones. Intra-divisional movements to find a return trip will not be considered as repositioning but as a directed return. Furthermore, when we talk about a repositioning zone, we mean a (re)loading zone induced by an empty trip.*

**Assumption 4** (Aggregate approach). *Our reference data are the OD matrices which are derived from annual surveys of individual carrier firms, and provide information on transported goods. The information provided by these matrices is an aggregation of all the information provided by each individual firm that has answered the survey. This leads us to consider an aggregated approach and thus to introduce a representative carrier firm, representing the diversity of companies. We will refer to this company as a carrier with a vehicle and who moves across the territory between loading and unloading zones. The company and its driver will be referred to interchangeably.*

**Assumption 5** (Probabilistic behaviour). *The carrier being a representative one intended to represent the diversity of carriers; this results in the randomness behavior due to the variability of the carriers. We will express probabilistic behavior by indicating a probabilistic choice.*

**Assumption 6** (Accessibility to market information). *The carrier is assumed to have knowledge on the freight market, of its flow asymmetries. This assumption follows from the aggregate approach.*

### 4.2 Methodology

Let us define  $\mathcal{L}$  the set of all the geographical zones of the considered territory (Assumption 1), denoted by

$$\mathcal{L} := \{1, 2, \dots, L\}. \quad (13)$$

Consider the long-haul and FTL transport service market (Assumptions 2) and a carrier under Assumption 4-5-6, having a single vehicle with capacity, denoted by  $\kappa$  such that  $\kappa \in \mathcal{K}$ , where

$$\mathcal{K} := \{0\} \cup [Q_{\min}, Q_{\max}], \quad (14)$$

denotes the set in which the vehicle capacity belongs. Assume that the carrier is based at the depot located at  $h \in \mathcal{L}$ .

Assume the carrier provides a shipment from its depot  $h$  to  $l \in \mathcal{L}$ . Assume also without loss of generality that the headhaul is defined as the direction from  $h$  to  $l$ . Hence, the carrier carries out a

fully loaded capacity shipment from the region of high demand  $h$ , that we call the loading zone, to the region of lower demand  $l$ , that we call the unloading zone, with  $h \neq l$ . The carrier is observed each time the vehicle arrives at a zone and at the end of each action; an action being either loading or unloading. So, the loading zone  $h$  is associated with a loading action characterized by a full load capacity  $Q$ , and is noted by  $(h, Q)$ . The unloading zone  $l$  is associated with an unloading action characterized by an empty vehicle's capacity, and is noted by  $(l, 0)$ . We now define the sets  $\mathcal{P}$  and  $\mathcal{D}$  of the loading (or pickup) and unloading (or delivery) zones respectively, namely

$$\mathcal{P} = \{(i, Q), i \in \mathcal{L}, Q \in [Q_{\min}, Q_{\max}]\}, \quad (15)$$

$$\mathcal{D} = \{(i, 0), i \in \mathcal{L}\}, \quad (16)$$

where  $\mathcal{L}$  is given by (13).

**Remark.** When we specify the action (loading or unloading) associated with a geographical location  $i \in \mathcal{L}$ , we will simply note  $i$  instead of  $(i, Q)$  or  $(i, 0)$  depending on the action. For example, we refer to the state  $(i, Q), i \in \mathcal{L}$  as the loading zone.

Due to the imbalance of trade and to secure its headhaul business, the carrier has to reposition its empty vehicle looking for a return shipment, from the unloading area  $(l, 0)$  to a (re)loading zone (or repositioning zone), say  $(k_1, Q)$  as depicted in Fig. 1. Under Assumption 3, note that the repositioning zone  $(k_1, Q)$  will be distinct of  $(h, Q)$  and  $(l, Q)$  which means that the carrier will not reposition the empty vehicle within  $l$  and  $h$ . From  $(k_1, Q)$  the carrier moves on to the unloading zone  $(h, 0)$  before completing their round-trip,  $(h, Q) - (l, 0) - (k_1, Q) - (h, 0)$  as displayed in Fig. 1.

Under Assumption 4, it is reasonable for the representative carrier to select not a single repositioning zone but a repositioning set consisting of a finite number of repositioning zones. This is because this carrier reflects the variability of all carriers and thus the variability of their repositioning choice behavior. As a result, the carrier behavior is viewed as probabilistic choice behavior. Fig. 1 shows an example of such a set consisting of three repositioning zones,  $\{(k_1, Q), (k_2, Q), (k_3, Q)\}$  where the carrier will have to reload goods. Given a choice of repositioning set, the carrier will again randomly select a repositioning zone within that set where they will move with their empty vehicle to reload the goods in order to finally unload them in the origin zone. This repositioning probability will be based on the Assumption 6.

The carrier's trip pattern is described by a trip-chaining starting and ending at the depot located in  $h \in \mathcal{L}$ . This pattern is induced by Assumption 2. Since the carrier has to load a full truck there is no other option than alternating between loading and unloading zones. The round-trip consists in four ordered and alternating loading and unloading zones as depicted in Fig. 1, such that

1. Initially, the carrier is located at the depot in  $h \in \mathcal{L}$  to carry out a loading action that is characterized by vehicle capacity  $Q$ . This initial node is displayed as  $(h, Q)$  in Fig. 1 to indicate that it is a loading node. The carrier travels with a fully loaded vehicle to the unloading node  $(l, 0)$ , with  $l \in \mathcal{L}$  and  $h \neq l$ ;
2. In  $(l, 0)$ , the goods are unloaded. The carrier selects according a repositioning probability one of the repositioning zones  $\{(k_1, Q), (k_2, Q), (k_3, Q)\}$ , say  $(k_1, Q)$  to move with an empty vehicle on to this area ;
3. In  $(k_1, Q)$ , the carrier has loaded new goods and move on with a fully loaded vehicle to the unloading node  $(h, 0)$  ;
4. In  $(h, 0)$ , the vehicle is empty. The carrier is located in the depot's area  $h$ . This is the end of the round-tour.

We are interested in the spatial repositioning behavior of the carrier in particular on the spatial probabilistic choice of a repositioning set. This choice is the result of a double decision-making process. The carrier is first confronted with the problem of choosing a repositioning set. It decides on a

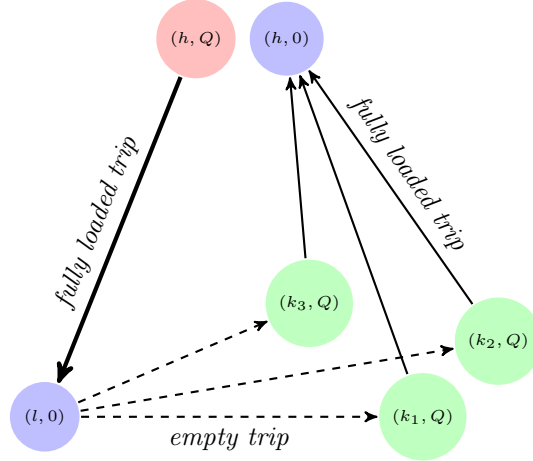


Fig. 1: Trip pattern and repositioning set with three repositioning zones  $(k_1, Q)$ ,  $(k_2, Q)$ ,  $(k_3, Q)$ .  $(h, Q) - (l, 0)$  is the headhaul direction with  $(h, Q)$  the loading node and  $(l, 0)$  the unloading node;  $(k_1, Q)$  to  $(h, 0)$  is one return shipment with  $(k_1, Q)$  a repositioning node and  $(l, 0)$  an unloading node. An example of round tour is  $(\ell, Q) - (u, 0) - (k_1, Q) - (\ell, 0)$ , for the sake of clarity since there is no ambiguity the round-tour will be denoted by  $hlk_1h$  since the trip is alternating and ordered.

repositioning set by arbitrating between different known alternatives. This repositioning set consists of a finite number of repositioning zones, and the decision is indicated by a probability of choice. Secondly, given a repositioning set, the carrier makes a decision-making rule within this set to make individual repositioning choices. Based on the knowledge of the market (Assumption *refassumption:knowledge of the market*) through the reloading opportunities which are characterized by the spatial distribution of reloading probabilities, the carrier assigns to each of the zones of the set a repositioning probability. The reloading probability is defined with respect to the headhaul direction, as the probability of finding a return shipment resulting from an empty repositioning trip. To investigate this problem we develop the two-stage solution scheme described in Section 5, as follow

### 1. Probabilistic repositioning set choice

The objective of the first step is to select a repositioning set. We investigate this choice as a result of a decision making process in which the carrier makes a trade-off between the different alternatives and selects one of them randomly under utility assumptions. The knowledge of the market through the spatial distribution of the reloading probability together with an empty distance criterion, influence the carrier's preferences and thus they can choose a repositioning set in a probabilistic way.

### 2. Probabilistic choice of unit repositioning

Given a repositioning set choice selected in step 1, the second step aims at choosing how the carrier will assign its individual repositioning decisions. Its individual repositioning choices are induced by market knowledge through the reloading probability distribution by evaluating a repositioning probability for each zone within the set (Assumption 6).

The technical issue is to determine the spatial reloading probability (Section 6). To do so, we introduce a random walk on a bipartite graph (Section 6.1). The transition matrix of the underlying Markov chain is empirically estimated by calibrating a Huff-type spatial interaction model on the basis of the OD matrices (Section 6.2). The definition and estimation of this probability is discussed in Section 6.3. From this spatial reloading probability, the repositioning probability within the repositioning set is easily determined.

## 5 Spatial probabilistic repositioning choice problem of the carrier

In this section, we address the problem of the carrier probabilistic choice of the spatial repositioning. The carrier is considered from the point of view of a consumer having to satisfy a need, namely, finding a return shipment resulting from a repositioning of the empty vehicle to secure the business. This choice is the result of a double decision-making process. At first, the carrier chooses a repositioning set consisting of a number  $n$  of repositioning zones by arbitrating between the different alternatives and selects one of them in a probabilistic way. The number  $n$  is assumed to be fixed by the carrier. In a second step, within the repositioning set, the carrier decides on the assignment of the individual repositioning zones by evaluating a repositioning probability under the knowledge of the spatial distribution of the reloading probability. We present the methodology in detail and refer to the next section for technical details.

To understand the repositioning behavior and decision-making process, it is necessary to know the carrier's preferences between the different alternatives. An alternative is described as a set of potential repositioning zones and are described using individual attributes defined for each zone. In the absence of data on the carrier's preferences, we first make assumptions on the attributes (Eqs. (17)-(31)), and then on the carrier's preferences (Eqs. (18)-(19)). As a result, the probability of choosing a repositioning set is derived (Eq. (20)). Based on the knowledge of the spatial distribution of the reloading probability (Eq. (31)), the carrier assigns a repositioning probability to each area of the repositioning set (Eq. 21). The main challenge is the computation of Eq. (31) which is postponed to Section 6.3.

The carrier is assumed to provide a shipment from the loading zone  $h$  to the unloading zone  $l$ . Before defining the individual attributes, let's define the set  $\mathcal{P}^{(hl)}$  consisting of all potential reloading zones excluding  $h$  and  $l$ , namely

$$\mathcal{P}^{(hl)} = \mathcal{P} \setminus \{(h, Q), (l, Q)\},$$

where  $\mathcal{P}$  is given by (15).

To measure the benefits and costs of an alternative, the carrier performs a trade-off to rationally assess the costs of the alternative against its benefits. The first important criteria we assume the carrier is focused on is the loaded run rate defined for a potential reloading zone  $k$ , with respect to the round-tour  $hlkh$ , i.e.,

$$r_{hlkh}^{load}(k) = 1 - \frac{d_{lk}}{d_{hlkh}}, \quad k \in \mathcal{P}^{(hl)}, \quad (17)$$

where  $d_{lk}$  is the empty distance trip, and  $d_{hlkh}$  is the total distance of the round-tour  $hlkh$ . It provides information on the percentage of loaded mileage (or empty mileage) covered relative to  $hlkh$  (or empty mileage and therefore not profitable).

The second criteria the carrier is focused on is the spatial reloading probability  $p_{hl}^{rel}(k)$  which measures the attractiveness of potential reloading zones. This probability will be given by (31) and gives information on reloading opportunities relative to the headhaul direction. Section 6 is devoted to the calculation of this probability.

Assume that the carrier specifies a number  $n$  of repositioning zones that constitute the choice of the repositioning set, with  $1 < n \leq L - 2$ . He determines the value of  $n$  based on the volume of the headhaul business and the knowledge of the market. We do not consider the following trivial case  $n = 1$ . In this case the problem would be reduced to a choice of a single repositioning zone that is incompatible with Assumption 4.

The carrier assigns certain constraints to each potential repositioning zone through two cut-off values to control both attributes (17) and (31). It will not select an area as an eligible repositioning zone :

1. if the loaded run rate (17) is less than the threshold value  $r$ . In other words, if the potential repositioning zone is considered too expensive (the empty run rate is too high or the loaded run rate is too low) by the carrier ;

2. if the probability (31) of reloading is below a threshold value  $p$ . In other words, if the potential relocation zone is not considered attractive enough in terms of reloading possibilities by the carrier.

We define the admissible set  $\mathcal{C}^{(hl)}$  with respect to the headhaul direction  $hl$ , the set of all potential repositioning zones satisfying minimum conditions characterized by the above constraints, namely,

$$\mathcal{C}^{(hl)} := \left\{ k \in \mathcal{P}^{(hl)} \mid r_{hlkh}^{load}(k) \geq r \text{ and } p_{hl}^{rel}(k) \geq p \right\}.$$

We define now the alternative set  $\mathcal{A}^{(hl)}$  with respect to  $hl$  as the collection of all the subsets consisting of  $n$  elements of the admissible set  $\mathcal{C}^{(hl)}$ .

The utility of any alternative  $R \in \mathcal{A}^{(hl)}$  such that all elements  $k$  of this alternative is characterized by the vector of attributes  $(r_{hlkh}^{load}(k), p_{hl}^{rel}(k))$  is defined by aggregate utility functions, namely

1. the perfect type utility function, i.e.,

$$u: \mathcal{A}^{(hl)} \rightarrow \mathbb{R}^+$$

$$R \mapsto u(R) = \sum_{k=1}^n r_{hlkh}^{load}(k) + p_{hl}^{rel}(k); \quad (18)$$

2. the Cobb-Douglas type utility function, i.e.,

$$v: \mathcal{A}^{(hl)} \rightarrow \mathbb{R}^+$$

$$R \mapsto v(R) = \sum_{k=1}^n r_{hlkh}^{load}(k) p_{hl}^{rel}(k). \quad (19)$$

**First stage decision-making process** The model is based on the premise of the probability that the carrier confronted with a set of alternatives will select a particular repositioning set is directly proportional to the perceived utility of each alternative. The choice behavior can be viewed as probabilistic. As a result, it is unlikely that any given alternative will be selected exclusively unless no other alternatives exist. The probability that the carrier will select the alternative  $R_u$  from  $\mathcal{A}^{(hl)}$ , with  $u = 1, \dots, |\mathcal{A}^{(hl)}|$  ( $|\mathcal{A}^{(hl)}|$  designates the cardinal of the set  $\mathcal{A}^{(hl)}$ ), corresponds to the utility of  $R_u$  renormalized by the sum of the utilities of alternatives, i.e.,

$$P(R_u) = \frac{U(R_u)}{\sum_{R_s \in \mathcal{A}^{(hl)}} U(R_s)}, \quad (20)$$

where  $U(\cdot)$  is either  $u(R_u)$  or  $v(R_u)$ .

**Second stage decision-making process** The carrier will then select a repositioning zone within  $R_u$  according the repositioning probability, for any repositioning zone  $k \in R_u$ ,

$$p_{hl}^{rep}(k) = \frac{p_{hl}^{rel}(k)}{\sum_{\ell \in R_u} p_{hl}^{rel}(\ell)}. \quad (21)$$

**Remark.** Under an assumption of rationality of the carrier the set repositioning choice with highest probability (20) coincides with the choice of the alternative associated with the highest level of utility. In that case, the first stage decision-making process would be defined deterministically. The carrier's decision problem therefore amounts to finding an optimal set  $R^* \in \mathcal{C}^{(hl)}$  maximizing an aggregate utility function ((18) or (19)). More formally, we are looking for  $R^* \in \mathcal{C}^{(hl)}$  such that

$$\max_{R \in \mathcal{C}^{(hl)}} \sum_{k=1}^n r_{hlkh}^{load}(k) + p_{hl}^{rel}(k) \left( \text{or } \max_{R \in \mathcal{C}^{(hl)}} r_{hlkh}^{load}(k) p_{hl}^{rel}(k) \right)$$

Note that solving the above maximization problem is elementary in this case. Since the utility function is additive with positive local utilities. The problem amounts to calculating the utilities of the  $(L-2)$  zones of  $\mathcal{L} \setminus \{h, l\}$ , to order them and take the  $n$  highest values and the associated zones.

## 6 The spatial reloading probability distribution

The objective of this section is to develop a methodology to access the spatial distribution of the reloading probability. We first introduce a random walk on a bipartite graph and the mathematical framework. The transition matrix of the underlying Markov chain is then empirically estimated by calibrating an extension of the Huff model on the basis of the OD matrices. The definition and estimation of the reloading probability is finally discussed.

### 6.1 Formulation of the carrier behavior with discrete Markov chain framework

As discussed in Section 4.2, the problem of evaluating the probability of spatial reloading is defined by means of a random walk on a bipartite graph. The idea is to represent the market dynamics through a random walker.

The network model we consider is a bipartite graph since the set of nodes can be partitioned into two disjoint sets  $\mathcal{P}$  and  $\mathcal{D}$ . The graph will be also labelled with non-negative values. Thus the network will be modelled by a bipartite graph with weighted edges, defined by

$$G = (\mathcal{P}, \mathcal{D}, E),$$

where,

- the set of loading nodes  $\mathcal{P}$  is given by (15) ;
- the set of unloading nodes  $\mathcal{D}$  is given by (16) ;
- $E = \left\{ \left( (i, Q), (j, 0) \right) \in \mathcal{P} \times \mathcal{D} \text{ or } \left( (i, 0), (j, Q) \right) \in \mathcal{D} \times \mathcal{P} \right\}$  is the set of arcs of the network.

The adjacency matrix  $\mathbf{W}$  associated with  $G$  is the  $2L \times 2L$  matrix of the form (1), namely

$$\mathbf{W} = \begin{pmatrix} \mathbf{0}_L & \mathbf{T} \\ \mathbf{T}^t & \mathbf{0}_L \end{pmatrix}, \quad (22)$$

where

- $\mathbf{0}_L$  is the null matrix of dimension  $L \times L$ , corresponding to the edges whose nodes are of the same kind ;
- $\mathbf{T}$  is the OD matrix (6) which connects the loading zones of  $\mathcal{P}$  to the unloading zones of  $\mathcal{D}$ ;  $\mathbf{T}^t$  designates the transpose of  $\mathbf{T}$  which link unloading nodes to loading nodes.

We use a random walk on  $G$  to define the spatial reloading probability. This dynamic corresponds to a bi-dimensional, homogeneous, and discrete time Markov chain  $X_n = \{(L_n, K_n)\}_{n \geq 0}$  with finite state space

$$\mathcal{S} := \mathcal{P} \cup \mathcal{D} = \{(i, \kappa), i \in \mathcal{L}, \kappa \in \mathcal{K}\},$$

where  $\mathcal{L}$  and  $\mathcal{K}$  are given respectively by (13) and (14). with initial distribution

$$\mathbb{P}((L_0, K_0) = (h, Q)) = 1.$$

The transition matrix which gives the probability that the Markov chain moves from one state (loading or unloading) to another one (unloading or loading) is denoted by

$$\mathbf{P} = \left( p_{(i, \kappa), (j, \kappa')} \right)_{(i, \kappa) \in \mathcal{S}, (j, \kappa') \in \mathcal{S}}.$$

The one-step transition on  $\mathcal{S}$  can be written in the general form, for  $(i, \kappa) \in \mathcal{S}$  and  $(j, \kappa') \in \mathcal{S}$

$$\mathbb{P}(L_{n+1} = j, K_{n+1} = \kappa' | L_n = i, K_n = \kappa) = p_{(i, \kappa), (j, \kappa')}, \quad (23)$$



where  $\sum_{(j,\kappa') \in \mathcal{S}} p_{(i,\kappa),(j,\kappa')} = 1$  and  $p_{(i,\kappa),(j,\kappa')} \geq 0$ . Following the general form of a transition matrix on a bipartite graph (4), (23) yields

$$\mathbf{P} = \begin{pmatrix} \mathbf{0}_L & \mathbf{P}^{(p)} \\ \mathbf{P}^{(d)} & \mathbf{0}_L \end{pmatrix}, \quad (24)$$

where  $\mathbf{0}_L$  is the  $L \times L$  dimensional null matrix corresponding to one-step transitions whose states are of the same kind.  $\mathbf{P}^{(p)}$  is the  $L \times L$  dimensional matrix corresponding to the one-step transition from a loading state to an unloading one. It characterizes fully loaded vehicle trips. Some restrictions are imposed on  $\mathbf{P}^{(p)}$ . As our study is based on an interregional OD matrix, this sub-matrix is with null diagonal. Let  $i \in \mathcal{L}$  be a loading state (or  $(i, Q) \in \mathcal{P}$ ). Following the definitions of transition probabilities (2) and (22), the conditional probabilities of unloading in  $j \in \mathcal{L}$ , given the loading zone in  $i$  are given by

$$p_{(i,Q),(j,0)} = \mathbb{P}(L_{n+1} = j, K_{n+1} = 0 | L_n = i, K_n = Q) = \begin{cases} \frac{T_{ij}}{\sum_{\substack{j \in \mathcal{L} \\ j \neq i}} T_{ij}} = F_{j|i} & \text{if } i \neq j \\ 0 & \text{if } i = j, \end{cases} \quad (25)$$

where  $F_{j|i}$  is given by (11).

The matrix  $\mathbf{P}^{(d)}$  of dimension  $L \times L$  corresponds to one-step transition from an unloading state to a loading one with the same restrictions on the diagonal as  $\mathbf{P}^{(p)}$ . It characterizes empty vehicle trips. Let  $j \in \mathcal{L}$  be an unloading state (i.e.,  $(j, 0) \in \mathcal{D}$ ). From (3) and (22), the conditional probabilities of loading in  $i \in \mathcal{L}$  given the unloading zone in  $j$  are given by

$$p_{(i,Q),(j,0)} = \mathbb{P}(L_{n+1} = i, K_{n+1} = Q | L_n = j, K_n = 0) = \begin{cases} \frac{T_{ij}}{\sum_{\substack{i \in \mathcal{L} \\ i \neq j}} T_{ij}} = F_{i|j}, & \text{if } i \neq j \\ 0 & \text{if } i = j, \end{cases} \quad (26)$$

where  $F_{i|j}$  is given by (12).

As stated in (25) and (26), the empirical counterparts of the transition probabilities can be easily derived from the OD matrices. Yet, in the following, we propose a mathematical modelling of those probabilities. This approach is justified by both practical and methodological arguments. First, the nature and quantity of the data do not allow an empirical estimation of the transition probabilities from the matrices. Second, this study aims at providing more visibility in the transportation process, thus an explanatory model for the estimation of  $\mathbf{P}^{(p)}$  and  $\mathbf{P}^{(d)}$  is called for.

## 6.2 Estimation of transition probabilities

We focus now on the empirical estimation of the transition matrix (24). Markov chains have been widely used to model the sequences of decisions and actions composing trip-chaining behavior (Thill et Thomas (1987), Thill et Timmermans (1992)). Some applications of Markov processes to trip-chaining behavior estimate the transition matrix by using random utility choice model (Ben-Akiva et al. (1978), Lerman (1979)) or a gravity-type model (Borgers et Timmermans (1986)). We propose a procedure for estimating the matrix using a Huff-type spatial interaction model. First, we present an extension of the Huff (5) to characterize a carrier's conditional choice of destinations; decisions that concern loading or unloading zones.

### 6.2.1 Formulation of the conditional Huff model

This choice of modelling is motivated by two main arguments. The first is that the carrier reasons sequentially at a fixed origin to determine the next destination; a conditional probability approach is therefore natural. The second is that origin areas have specific economic and structural characteristics,

so it seems more relevant to consider (5) at an origin level. This guarantees greater flexibility and a better adjustment of the model. These considerations led us to pose  $L$  Huff models on the conditional unloading (resp. loading) distribution given that each loading (resp. unloading) zone. The set of these sub-models is thus parametrized by  $L$  vectors  $(\alpha_i^{(p)}, \beta_i^{(p)})_{i=1,\dots,L}$  (resp.  $(\alpha_i^{(d)}, \beta_i^{(d)})_{i=1,\dots,L}$ ), each being estimated locally on a subset of the OD matrix  $\mathbf{T}$  (resp.  $\mathbf{T}^t$ ).

**Unloading conditional distribution given loading zone.** Starting from a loading zone  $i$  which is characterized by its economic weight  $E_i$  (Eq. (9)), we define the utility function of the unloading zone  $j$ , denoted  $U_{ij}^{(p)}$  for the carrier located at  $i$  as follows

$$U_{ij}^{(p)} = R_j \alpha_i^{(p)} / \bar{C}_{ij}^{\beta_i^{(p)}},$$

where  $R_j$  is the reception rate (Eq. 10) which reflects the attractiveness of the unloading state  $j$ ;  $\bar{C}_{ij}$  the mean transportation cost between the centroids of  $i$  and  $j$  depending on the distance  $d_{ij}$  (See Appendix D). For a given loading zone  $i$ ,  $\alpha_i^{(p)}$  and  $\beta_i^{(p)}$  represent the conditional sensitivity of the utility of the unloading zone  $j$ , as a result of a change in  $R_j$  and  $\bar{C}_{ij}$  respectively. In particular, the parameter  $\alpha_i^{(p)}$  measures elasticity of the attractiveness relative to the loading state  $i$ . Thus, an increase in reception  $R_j$  will lead to greater increase in utility the higher the economic weight  $E_i$  of the loading location. The higher the reception rate  $R_j$ , the higher the utility; and the higher the transport cost between  $i$  and  $j$  the lower the utility, hence for all loading zones  $i \in \mathcal{L}$ ,  $\alpha_i^{(p)} > 0$  and  $\beta_i^{(p)} > 0$ . We observe a strong positive correlation between the  $\{\hat{\alpha}_i^{(p)}, i \in \mathcal{L}\}$  and  $\{E_i, i \in \mathcal{L}\}$ , where  $\hat{\alpha}_i^{(p)}$  designates the estimator of  $\alpha_i^{(p)}$  (See Tab. 7 in Appendix B). This is because the parameter  $\alpha_i^{(p)}$ ,  $i = 1, \dots, L$ , represents the size effect associated with a loading zone  $i$ , and captures information about the economic weight of  $i$ .

The probability of an unloading destination  $j \in \mathcal{L}$ , such that  $j \neq i$  given the loading zone  $i$  is defined as the ratio of the utility  $U_{ij}^{(p)}$  with the sum of the utilities of all the unloading destination alternatives

$$\begin{aligned} p_{(i,Q),(j,0)} &= \mathbb{P}((L_{n+1} = j, K_{n+1} = 0) \mid (L_n = i, K_n = Q)), \quad j \neq i \\ &= \frac{U_{ij}^{(p)}}{\sum_{\substack{k \in \mathcal{L} \\ k \neq i}} U_{ik}^{(p)}} = \frac{R_j \alpha_i^{(p)} / \bar{C}_{ij}^{\beta_i^{(p)}}}{\sum_{\substack{k \in \mathcal{L} \\ k \neq i}} R_k \alpha_i^{(p)} / \bar{C}_{ik}^{\beta_i^{(p)}}}, \end{aligned} \quad (27)$$

where  $p_{(i,Q),(j,0)}$  is the conditional probability of unloading in  $j$  given that the loading in  $i$ , with  $\sum_{j \neq i} p_{(i,Q),(j,0)} = 1$ , and  $0 < p_{(i,Q),(j,0)} < 1$ . The choice of the unloading destination results from drawing under the probabilities given in Eq. (27).

For any loading zone  $i \in \mathcal{L}$ , the unknown parameter  $\theta_i^{(p)} := (\alpha_i^{(p)}, \beta_i^{(p)})$  will be estimated on the basis of the OD matrix  $\mathbf{T}$  (6). The statistical estimation procedure is postponed to Appendix B. The estimator of  $p_{(i,Q),(j,0)}$  yields

$$\hat{p}_{(i,Q),(j,0)} = \frac{\hat{U}_{ij}^{(p)}}{\sum_{\substack{k \in \mathcal{L} \\ k \neq i}} \hat{U}_{ik}^{(p)}} = \frac{R_j^{\hat{\alpha}_i^{(p)}} / \bar{C}_{ij}^{\hat{\beta}_i^{(p)}}}{\sum_{\substack{k \in \mathcal{L} \\ k \neq i}} R_k^{\hat{\alpha}_i^{(p)}} / \bar{C}_{ik}^{\hat{\beta}_i^{(p)}}}. \quad (28)$$

where, for any loading state  $i$ ,  $\hat{\theta}_i^{(p)} := (\hat{\alpha}_i^{(p)}, \hat{\beta}_i^{(p)})$  designates the estimator of  $\theta_i^{(p)}$ .

**Loading conditional distribution given unloading zone.** In the same way as the previous paragraph, the utility function of the loading destination zone  $j$  starting from a given unloading origin zone  $i$  which is characterized by its economic weight  $E_j$  (Eq. 9) is defined as follows

$$U_{ij}^{(d)} = E_j \alpha_i^{(d)} / \bar{C}_{ij}^{\beta_i^{(d)}},$$

where  $U_{ij}^{(d)}$  is the utility function of the loading location  $j$  for a carrier originating from the unloading state  $i$ ;  $E_j$  characterizes the attractiveness of the loading zone destination  $j$ , measured by its emission rate (9);  $\bar{C}_{ij}$  is the mean transport cost from the centroid of the unloading zone  $i$  to the centroid of the loading zone  $j$  (Appendix D). For a given unloading zone  $i$ ,  $\alpha_i^{(d)}$  and  $\beta_i^{(d)}$  represent the conditional elasticities of the attractiveness and the transportation cost from  $i$  to  $j$ . The higher the emission rate  $E_j$ , the higher the utility; and the higher the transportation cost from  $i$  to  $j$  the lower the utility, thus for all origin unloading zone  $i$ ,  $\alpha_i^{(d)} > 0$  and  $\beta_i^{(d)} > 0$ . The parameter  $\alpha_i^{(d)}$  characterizes the size effect of the unloading zone  $i$  (with a strong positive correlation between  $\alpha_i^{(d)}$  and  $R_i$ , see Appendix C).

The conditional probability of loading in  $j$  given the unloading area  $i$ ,  $p_{(i,0),(j,Q)}$ , is equal to the utility of the loading location  $j$  divided by the total sum of the utilities of the potential  $(L-1)$  loading locations. The estimator of the conditional loading distribution given the unloading at  $i \in \mathcal{L}$  is given by

$$\begin{aligned} \hat{p}_{(i,0),(j,Q)} &= \hat{\mathbb{P}}((L_{n+1} = j, K_{n+1} = Q) | (L_n = i, K_n = 0)), \quad j \in \mathcal{L} \setminus \{i\} \\ &= \frac{\hat{U}_{ij}^{(d)}}{\sum_{\substack{k \in \mathcal{L} \\ k \neq i}} \hat{U}_{ik}^{(d)}} = \frac{E_j^{\hat{\alpha}_i^{(d)}} / \bar{C}_{ij}^{\hat{\beta}_i^{(d)}}}{\sum_{\substack{k \in \mathcal{L} \\ k \neq i}} E_k^{\hat{\alpha}_i^{(d)}} / \bar{C}_{ik}^{\hat{\beta}_i^{(d)}}}, \end{aligned} \quad (29)$$

with  $\hat{\theta}_i^{(d)} := (\hat{\alpha}_i^{(d)}, \hat{\beta}_i^{(d)})$  is the estimator of the unknown parameters  $\theta_i^{(d)} := (\alpha_i^{(p)}, \beta_i^{(d)})$  obtained by the estimation procedure described in Appendix C.

### 6.3 Definition of the spatial reloading probability distribution

Recall that  $\mathcal{P}^{(hl)}$  is the set of potential reloading candidates resulting from an empty trip from  $(l, 0)$ , and privates of both loading zones  $(ih, Q)$  and  $(l, Q)$ . We are interested in the probability of reloading in  $k \in \mathcal{P}^{(hl)}$  starting from the unloading state  $l$  to return to unload in  $h$ , denoted  $\mathbb{P}_{hl}^{rel}(k)$ . Due to the joint nature of a transport, the carrier has to return at  $h$  to complete the round-tour by repositioning the empty vehicle in  $k \in \mathcal{L}_{hl}^Q$ , which yields for all  $k \in \mathcal{P}^{(hl)}$

$$p_{hl}^{rel}(k) = \mathbb{P}(\{L_2 = k, K_2 = Q\} | \{L_0 = h, K_0 = Q\}, \{L_1 = l, K_1 = 0\}, \{L_3 = h, K_3 = 0\})$$

The above Equation yields

$$p_{hl}^{rel}(k) = \frac{\mathbb{P}(\{L_0 = h, K_0 = Q\}, \{L_1 = l, K_1 = 0\}, \{L_2 = k, K_2 = Q\}, \{L_3 = h, K_3 = 0\})}{\mathbb{P}(\{L_0 = h, K_0 = Q\}, \{L_1 = l, K_1 = 0\}, \{L_3 = h, K_3 = 0\})} \quad (30)$$

The numerator of (30) is calculated using the Markov property, hence

$$\begin{aligned} &\mathbb{P}(\{L_0 = h, K_0 = Q\}, \{L_1 = l, K_1 = 0\}, \{L_2 = k, K_2 = Q\}, \{L_3 = h, K_3 = 0\}) \\ &= p_{(h,Q),(l,0)} \times p_{(l,0),(k,Q)} \times p_{(k,Q),(h,0)} \end{aligned}$$

The computation of the denominator of (30) yields

$$\begin{aligned} &\mathbb{P}(\{L_0 = h, K_0 = Q\}, \{L_1 = l, K_1 = 0\}, \{L_3 = h, K_3 = 0\}) \\ &= \sum_{(z,Q) \in \mathcal{P}} \mathbb{P}(\{L_0 = h, K_0 = Q\}, \{L_1 = j, K_1 = 0\}, \{L_2 = z, K_2 = Q\}, \{L_3 = h, K_3 = 0\}) \\ &= p_{(h,Q),(l,0)} \sum_{(z,Q) \in \mathcal{P}} p_{(l,0),(z,Q)} p_{(z,Q),(h,0)} = p_{(h,Q),(l,0)} p_{((l,0),(h,0))}^2 \end{aligned}$$

Finally, the reloading probability is

$$p_{hl}^{rel}(k) = \frac{p_{((l,0),(k,Q))} p_{((k,Q),(h,0))}}{p_{((l,0),(h,0))}^2}, \quad k \in \mathcal{P}^{(hl)}, \quad (31)$$

where the quantity  $p_{((l,0),(h,0))}^2$  which designates the probability of unloading in  $h$  knowing that the carrier has unloaded in  $l$  in two steps is strictly positive, because the graph is bipartite and therefore periodic of period 2,  $p_{((l,0),(k,Q))}$  is the probability of reloading in  $k$  given the carrier unloaded its goods in  $l$ ;  $p_{((k,Q),(h,0))}$  represents the probability that the carrier unloads in  $i$  given goods have been loaded in  $k$ .

An estimator of (31) is given by

$$\hat{p}_{hl}^{rel}(k) = \frac{\hat{p}_{((l,0),(k,Q))}\hat{p}_{((k,Q),(h,0))}}{\hat{p}_{((l,0),(h,0))}^2}, \quad k \in \mathcal{P}^{(hl)}.$$

where  $\hat{p}_{((k,Q),(h,0))}$  is given by Eq. (28),  $\hat{p}_{((l,0),(k,Q))}$  by Eq. (29), and  $\hat{p}_{((k,Q),(i,0))}$ ,  $\hat{p}_{((l,0),(h,0))}^2$  is given by

$$\hat{p}_{((l,0),(h,0))}^2 = \sum_{(z,Q) \in \mathcal{P}} \hat{p}_{((l,0),(z,Q))}\hat{p}_{((z,Q),(i,0))}.$$

## 7 Numerical experiments

This section presents numerical experiments of the proposed methodology for the probabilistic repositioning choice problem. In the road freight transport sector, which does not fully benefit from new technologies to facilitate information sharing and feedback, data concerning carriers' repositioning zones play a major role in the construction of a transport price and are difficult to access. The objective is to have a better knowledge of the market and a better visibility of carriers' choices in terms of repositioning zones in order to shed light on more practical problems. These numerical simulations aim to highlight information from our model in order to illustrate how it could be used in practice. We illustrate this methodology in the context of Belgium. From a technical point of view, note that the spatial reloading probability underlying these numerical experiments focuses on the minimum of the Hellinger divergence estimate. We refer the reader to a comparison of estimators under various divergences for example in Appendices B and C. Tab. 1 lists the parameters used in these simulations. The parameters of the transportation cost are listed in Appendix D.

Parameters	Definition	Values
$n$	number of repositioning zones	3
$r$	loading run rate cut-off	0.65
$p$	reloading probability cut-off	0.08

Tab. 1: Simulation parameters

### 7.1 Empirical illustrations of the carrier's choice of repositioning set

In this section, we present some of the results of the proposed model to help practitioners in transport management. Fig. 2a represents the spatial distribution with respect to the headhaul direction (from Antwerp to Luxembourg depicted by the black arrow) of the reloading probability across Belgium. This probability (Eq. (31)) is estimated according to our methodology presented in Section 6.3. It reflects the carriers' reloading opportunities from the unloading area assumed to be of lower demand (the arrowhead in Fig. 2a) to a different zone. The value of these probabilities is represented by a colour gradient. The darker the colour, the higher the probability of reloading. From Fig. 2a, the three highest spatial reloading probabilities relative to the headhaul occur in provinces of Liège, East Flanders and Limburg, with

1. a probability of 0.21 for Liège ;
2. a probability of 0.19 for East Flanders ;
3. a probability of 0.18 for Limburg.

These values are given in Tab. 2. The first probability means that, given that the carrier has carried out a transport from Antwerp to Luxembourg, a carrier has a 21% chance of reloading goods in Liège from Luxembourg to finally unload in his place of origin (Antwerp). The carrier will thus complete the round trip (Antwerp-Luxembourg-Liège-Antwerp), provided that it decides to reposition the vehicle there. Similarly, these results also show that the carrier has a 19% chance of reloading goods in East Flanders and 18% chance of reloading in Limburg to return to Antwerp in order to unload the goods, provided it decides to reposition there.

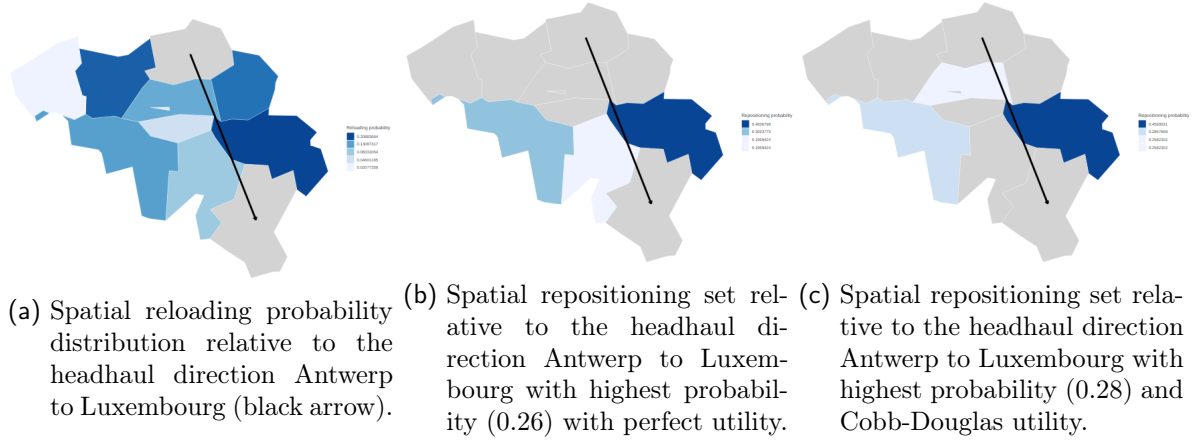


Fig. 2: Mapping of repositioning probabilities

loading zones	unloading zones	reloading zones	loaded run rates	reloading probabilities
Antwerp	Luxembourg	Brussels	0.67	0.03
Antwerp	Luxembourg	Limburg	0.65	0.18
Antwerp	Luxembourg	East Flanders	0.58	0.19
Antwerp	Luxembourg	Flemish Brabant	0.66	0.12
Antwerp	Luxembourg	West Flanders	0.58	0.02
Antwerp	Luxembourg	Walloon Brabant	0.71	0.05
Antwerp	Luxembourg	Hainaut	0.68	0.13
Antwerp	Luxembourg	Liège	0.80	0.21
Antwerp	Luxembourg	Namur	0.81	0.08

Tab. 2: Attributes' values for each zones except the headhaul direction: the loaded run rate and the reloading probabilities.

Tab. 2 lists all the potential reloading zones (before the carrier decision). Each of these zones is characterised by its two attributes: the reloading probability, and the loaded run rate with respect to its corresponding round trip. A reloading in Liège leads to the round tour Antwerp-Luxembourg-Liège-Antwerp which is characterised by the highest loaded run rate of 80%, or in other words by the lower empty run rate, i.e., 20%. This means that on the whole round-tour Antwerp-Luxembourg-Liège-Antwerp round, the carrier would make 20% of their journey empty. A reloading in East Flanders leads to the round-trip Antwerp-Luxembourg-East Flanders-Antwerp which is characterised by the lower loaded run rate, namely, 58%, or the highest empty rate, i.e., 42%.

The model provides additional information listed in Tab. 3, namely the empty trip distance (between centroids, under Assumption 1) and the mean of the transport cost distribution presented in Appendix D. These variables are the determinants of the associated Huff model for the conditional law of loading given unloading (Eq. 28). They can complement the information in Tab. 2 by giving an approximation of the hidden costs of repositioning. Let's take the example of a reloading in Liège, the distance travelled empty is 87 km, the associated cost which is carried out by the carrier is estimated at 235 € on average.

loading zones	unloading zones	reloading zones	empty distance (km)	average cost (€)
Antwerp	Luxembourg	Brussels	146.01	301.66
Antwerp	Luxembourg	Limburg	156.17	313.14
Antwerp	Luxembourg	East Flanders	213.02	377.37
Antwerp	Luxembourg	Flemish Brabant	150.27	306.47
Antwerp	Luxembourg	West Flanders	267.88	439.36
Antwerp	Luxembourg	Walloon Brabant	125.81	278.83
Antwerp	Luxembourg	Hainaut	165.66	323.86
Antwerp	Luxembourg	Liège	87.17	235.17
Antwerp	Luxembourg	Namur	86.30	234.19

Tab. 3: Empty distances with associated mean transportation cost

The spatial distribution of the reloading probability with respect to a headhaul direction does not necessarily reflect the carrier’s actual choice of relocation, but rather a state of the market. The carrier uses this information to make decisions and assess the spatial repositioning probability. To decide on a repositioning set, the carrier is faced with the problem of choosing a repositioning set consisting of an a priori fixed number of potential repositioning zones. This set is selected with probability (20). Figs. 2b and 2c represent the carrier choice of a repositioning set with three zones highest choice probability with respectively (18) and (19) utilities. After choosing the repositioning set, the carrier assesses one particular zone with the spatial repositioning probability. Tabs. 4 and 5 provide the spatial repositioning probability (21).

We also observe different behaviour depending on the utility function. Under the Cobb-Douglas utility function (Eq. 19), the carrier is willing to travel additional empty kilometres to have more chances to reload. Indeed, he is willing to go to Flemish Brabant, thus taking on average 306 € to cover the 150 empty repositioning kilometres, against repositioning in Namur for an average costs of 234 € for the perfect utility function (Eq. 18). Combining the information in Tabs. 3 with 4-5, we derive the average cost of repositioning relative to its repositioning set relative to (18)-(19). The repositioning set (Tab 4) is associated with the hidden average repositioning cost: 262.41 €. The repositioning set (Tab 5) is associated with the hidden average repositioning cost: 278.93 €.

loading zones	unloadings zones	repositioning zones	repositioning probabilities
Antwerp	Luxembourg	Hainaut	0.31
Antwerp	Luxembourg	Liège	0.49
Antwerp	Luxembourg	Namur	0.20

Tab. 4: Repositioning set choice with highest probability with perfect utility

loading zones	unloading zones	repositioning zones	repositioning probabilities
Antwerp	Luxembourg	Flemish Brabant	0.26
Antwerp	Luxembourg	Hainaut	0.29
Antwerp	Luxembourg	Liège	0.46

Tab. 5: Repositioning set choice with highest probability with Cobb-Douglas utility

## 8 Conclusion and perspectives

This study tackles the untreated, yet critical issue of formalizing the spatial choices of a carrier dealing with repositioning. It was motivated by the problems encountered by most actors in the freight transport industry, namely the lack of visibility on the repositioning choices of carriers. This question is a major blocking point to the development of transparency and visibility on the market.



Indeed, among the numerous opacities on the transport market, the behavior of its actors is particularly illegible, especially concerning the carriers' choices and strategies in terms of finding return shipments. Indeed, in a trade imbalance context, repositioning their vehicle to sustain their business is a necessity. Those behaviors reflecting the carriers own market knowledge and personal strategies are often not formalized by the carriers themselves and unknown to the shippers or freight forwarders, thus resulting in an impossibility to manage their costs and transport plans optimally. This question is challenging on multiple levels. On an operational level, providing explanatory tools to understand the decision-making process of a carrier on repositioning zones selection will remove a significant obstacle to information-sharing and visibility in the transport industry. On a technical one, it implies the development of an original model as it has not been studied.

The solution we propose is an original combination of statistical and econometric tools to model the opaque process of repositioning choices. The formal framework borrows from the economic theory of the consumer: we model the movements of a representative carrier, in the sense of a production unit. We adopt an aggregate approach based on the exploitation of national OD matrices as market representation. Thus the carrier is assumed to know the market and his movements represent its trends. There are three main modeling tools:

1. A random walk on a bipartite graph representing the flows exchanged between regions.
2. An adaptation of Huff's spatial interaction model to express the spatial reloading distribution from a given origin (with a loading or unloading action) on the national territory.
3. A choice procedure based on assumptions on the carrier's preferences.

The preliminary step consists in estimating the loading and unloading distributions on the territory. Huff's spatial interaction model is adapted locally to define on one hand the conditional distributions of unloadings given each origin location and on the other hand the distributions of loadings given each destination location. Those probabilities define the transition matrix of the Markov chain modeling the carrier spatial repositioning behavior. They are used to express two key notions in the choice procedure: the reloading probabilities of a repositioning candidate given an initial expedition from an origin  $h$  to a destination  $l$  and the repositioning probabilities among a selected set of candidates which are derived from the latter.

The proposed procedure is a two-steps decision-making process: First the carrier decides on the set of repositioning candidates through a trade-off between the reloading probabilities as an indicator of the locations' attractivity and the additional cost the supplementary trip induces. This trade-off is determined by the preferences of the carrier represented by utility functions. Secondly, within the selected set, we define the distribution of the repositioning probabilities from the reloading probabilities and draw a repositioning location.

Our methodology is illustrated from an operational standpoint by means of numerical simulations. They aim at providing transparency on the spatial behavior of carriers through indicators built from the different information contained in the underlying models.

## Research perspectives

We propose research perspectives to address the problem of modeling freight demand and the effect of trade imbalance on transportation price. The method we have developed can be used as a basic theoretical framework for modeling the effect of trade imbalance on the price of transportation. Indeed, one possible interpretation of the effect of trade imbalance on price is the notion of repositioning cost corresponding to the additional costs due to empty miles weighted by the risk for the carrier of not finding return freight. We have proposed an extension of the Huff spatial interaction model to express the spatial distribution of a load (resp. unloading) conditional on an unloading (resp. loading) on the national territory. The first simulations performed on the Belgian territory suggest that this local estimation procedure is reasonable and can be extended to the so-called four-stage models with the first stage called Generation and the second called Distribution. From a software development point of view, our methodology can be used in decision-making tools.

## A Statistical estimation based on $\phi$ -divergence

Divergences are widely used in statistics. We are concerned with finite discrete support distributions. For these types of distributions, [Liese et Vajda \(1987\)](#) and [Morales et al. \(1995\)](#) introduced the so-called minimum  $\phi$ -divergence estimators. The use of  $\phi$ -divergences criteria in statistics cover and improve classical methods in some cases. In particular, [Lindsay \(1994\)](#) and [Jiménez et Shao \(2001\)](#) show that the minimum Hellinger estimator is better than all other power divergence estimators in terms of efficiency and second order robustness. The quality of real observations motivates the use of robust estimators. In the case where the densities of the model are continuous, [Toma et Broniatowski \(2011\)](#) uses an approach based on the influence of the efficiency and the second-order robustness.

The estimation problems through  $\phi$ -divergences that we are considering can be described as follows. Let  $\{P_\theta, \theta \in \Theta\}$  be an identifiable model with  $\Theta$  an open subset of  $\mathbb{R}^d$ , with  $d \geq 1$ . Consider the problem of estimating the unknown true value of the parameter  $\theta^*$  on the basis of an independent and identically distributed sample (i.i.d),  $X_1, \dots, X_n$  with probability measure (p.m)  $P_{\theta^*}$ . When all the p.m.  $P_\theta$  share the same discrete finite support  $\mathcal{S}$  which is independent of the parameter  $\theta$ , the  $\phi$ -divergence between  $P_\theta$  and  $P_{\theta^*}$  is defined by

$$\phi(P_\theta, P_{\theta^*}) = \sum_{j \in \mathcal{S}} \varphi \left( \frac{P_\theta(j)}{P_{\theta^*}(j)} \right) P_{\theta^*}(j),$$

where  $\varphi$  is a non-negative convex function defined on  $[0, +\infty[$  onto  $[0, +\infty]$  and satisfying  $\varphi(1) = 0$ .

The minimum  $\phi$ -divergence estimators  $\hat{\theta}_n$  for the parameter  $\theta^*$  is defined as the solution of the minimization problem

$$\hat{\theta}_n := \operatorname{argmin}_{\theta \in \Theta} \phi(P_\theta, P_n) = \operatorname{argmin}_{\theta \in \Theta} \sum_{j \in \mathcal{S}} \varphi \left( \frac{P_\theta(j)}{P_n(j)} \right) P_n(j),$$

with  $\phi(P_\theta, P_n)$  is the plug-in estimator of  $\phi(P_\theta, P_{\theta^*})$ , and  $P_n := \frac{1}{n} \sum_{i=1}^n \delta_{X_i}$  is the empirical measure associated with the sample, where  $\delta_x$  designates the Dirac measure at  $x$ .

## B Estimation of the conditional unloading distribution given the loading

For any given loading state  $i \in \mathcal{L}$ , assume we observe an i.i.d sample  $(X_\ell^{(i)}, \ell \in \mathcal{L} \setminus \{i\})$  with common and unknown conditional probability distribution  $P_{\theta_i^{*(p)}}$ . Assume also that  $\theta_i^{*(p)} \in \Theta_i^{(p)}$ , an open subset of  $\mathbb{R}_+ \times \mathbb{R}_+$ . For any given loading state  $i$ , let's denotes by  $P_i(j)$  the empirical and conditional probability distribution for these samples defined by Eq. (11), namely

$$P_i(j) = F_{j|i} = T_{ij}/T_{i\bullet}, \quad j \in \mathcal{L} \setminus \{i\}. \quad (32)$$

Assume that the distribution  $P_{\theta_i^{*(p)}}$  is defined by the proposed Huff model (27). The objective is to estimate for each given loading state  $i \in \mathcal{L}$ , the true unknown parameter  $\theta_i^{*(p)}$  using  $\phi$ -divergences, on the basis of the i.i.d. sample  $(X_\ell^{(i)}, \ell \in \mathcal{L} \setminus \{i\})$ . Define the parametric model, for any given loading state  $i \in \mathcal{L}$

$$\mathcal{M}_i^{(p)} := \left\{ P_{\theta_i^{(p)}}, \theta_i^{(p)} \in \Theta_i^{(p)} \right\}, \quad (33)$$

where

$$\Theta_i^{(p)} := \left\{ \theta_i^{(p)} = (\alpha_i^{(p)}, \beta_i^{(p)}), \alpha_i^{(p)} \subseteq \mathbb{R}_+, \beta_i^{(p)} \subseteq \mathbb{R}_+ \right\}, \quad (34)$$

$$P_{\theta_i^{(p)}}(j) = \frac{U_{ij}^{(p)}}{\sum_{\substack{k \in \mathcal{L} \\ k \neq i}} U_{ik}^{(p)}} = \frac{R_j^{\alpha_i^{(p)}} / C_{ij}^{\beta_i^{(p)}}}{\sum_{\substack{k \in \mathcal{L} \\ k \neq i}} R_k^{\alpha_i^{(p)}} / C_{ik}^{\beta_i^{(p)}}}, \quad j \in \mathcal{L} \setminus \{i\}. \quad (35)$$

For a given loading state  $i \in \mathcal{L}$ , all the distributions  $P_{\theta_i^{(p)}}$  have the same discrete and finite support which does not depend on  $\theta_i^{(p)}$ . The model (33)-(34)-(35) being identifiable, the minimum  $\phi$ -divergence estimators of  $\theta_i^{(p)}$ , denoted by  $\hat{\theta}_{i,\varphi}^{(p)}$  is defined as the unique solution of the following minimization problem :

$$\hat{\theta}_{i,\varphi}^{(p)} = \underset{\theta_i^{(p)} \in \Theta_i^{(p)}}{\operatorname{argmin}} \phi \left( P_{\theta_i^{(p)}}, P_i \right) = \underset{\theta_i^{(p)} \in \Theta_i^{(p)}}{\operatorname{argmin}} \sum_{\substack{j \in \mathcal{L} \\ j \neq i}} \varphi \left( \frac{P_{\theta_i^{(p)}}(j)}{P_i(j)} \right) P_i(j),$$

where  $\phi \left( P_{\theta_i^{(p)}}, P_{(L-1),i} \right)$  is the plug-in estimators of  $\phi \left( P_{\theta_i^{(p)}}, P_{\theta_i^{*(p)}} \right)$ , with  $P_i(j)$  given by (32).

In the following we consider three divergences :

1. modified Kullback-Leibler divergence ;
2. Hellinger divergence ;
3. Pearson's  $\chi^2$  divergence.

**Maximum likelihood Estimator.** For any loading state  $i \in \mathcal{L}$ , the maximum likelihood estimator of  $\theta_i^{(p)}$  is associated with the modified Kullback-Leibler divergence, denoted  $KL_m$ , and defined by the convex function  $\varphi(x) = -\log(x) + x - 1$ . The maximum likelihood estimator,  $\hat{\theta}_{i,ML}^{(p)}$  is the unique solution of the equation, for all given loading state  $i$

$$\hat{\theta}_{i,ML}^{(p)} = \underset{\theta_i^{(p)} \in \Theta_i^{(p)}}{\operatorname{argmax}} \sum_{\substack{j \in \mathcal{L} \\ j \neq i}} \log \left( P_{\theta_i^{(p)}}(j) \right) P_i(j).$$

**The minimum of the Hellinger divergence estimator.** The Hellinger  $\phi$ -divergence, denoted  $H$  is associated with the convex function  $\varphi(x) = 2(\sqrt{x} - 1)^2$ . For any loading state  $i \in \mathcal{L}$ , the minimum of the Hellinger divergence estimator of  $\theta_i^{(p)}$ , noted  $\hat{\theta}_{i,H}^{(p)}$  is the unique solution of the equation

$$\hat{\theta}_{i,H}^{(p)} = \underset{\theta_i^{(p)} \in \Theta_i^{(p)}}{\operatorname{argmin}} \sum_{\substack{j \in \mathcal{L} \\ j \neq i}} \left( \sqrt{\frac{P_{\theta_i^{(p)}}(j)}{P_i(j)}} - 1 \right)^2 P_i(j).$$

**The minimum of Pearson's  $\chi^2$  divergence estimator.** The Pearson's  $\chi^2$  divergence is associated with the convex function  $\varphi(x) = \frac{1}{2}(x - 1)^2$ . For any loading state  $i \in \mathcal{L}$ , the minimum the Pearson's  $\chi^2$  divergence noted  $\hat{\theta}_{i,\chi^2}^{(p)}$  is the unique solution of the equation

$$\hat{\theta}_{i,\chi^2}^{(p)} = \underset{\theta_i^{(p)} \in \Theta_i^{(p)}}{\operatorname{argmin}} \sum_{\substack{j \in \mathcal{L} \\ j \neq i}} \frac{\left( P_{\theta_i^{(p)}}(j) - P_i(j) \right)^2}{P_i(j)}.$$

## Estimation results

Tab. 6 shows that for all loading zones, the estimators are interior points of the optimization domain (except for Luxembourg (BE34) and its  $\beta$  estimator), and that the estimators are very close. The performance in terms of MSE is also of the same order as shown in Tabs. 6 and 8. Moreover, Tab. 7 shows that the vectors of  $\alpha_i$  estimators and of  $E_i$  emission rates are significantly and positively correlated, and that there is no significant difference between the three criteria. The Huff model that we propose recovers a large part of the economic weight of the loading zone.

NUTS	Names	$KL_m$		$H$		$\chi^2$		$KL_m$	$H$ $MSE_i$	$\chi^2$
		$\alpha_i$	$\beta_i$	$\alpha_i$	$\beta_i$	$\alpha_i$	$\beta_i$			
BE10	Brussels	0.30	1.38	0.29	1.40	0.27	1.47	0.00106	0.00108	0.00118
BE21	Antwerp	1.34	1.28	1.38	1.36	1.48	1.51	0.00581	0.00577	0.00573
BE22	Limburg	1.16	2.39	1.21	2.44	1.32	2.58	0.00654	0.00646	0.00627
BE23	East Flanders	1.76	0.54	1.79	0.62	1.87	0.70	0.01200	0.01180	0.01170
BE24	Flemish Brabant	0.93	0.89	0.93	0.92	0.95	0.99	0.00536	0.00548	0.00589
BE25	West Flanders	1.34	1.04	1.41	1.05	1.58	1.17	0.00850	0.00848	0.00813
BE31	Walloon Brabant	0.32	0.64	0.32	0.70	0.28	0.92	0.00346	0.00353	0.00396
BE32	Hainaut	1.01	0.56	1.03	0.61	1.12	0.83	0.00684	0.00690	0.00723
BE33	Liège	0.87	2.50	0.94	2.55	1.14	2.69	0.00353	0.00356	0.00366
BE34	Luxembourg	1.01	0.10	1.01	0.10	1.01	0.10	0.00294	0.00294	0.00294
BE35	Namur	0.84	3.28	0.90	3.40	0.98	3.68	0.00601	0.00619	0.00664

Tab. 6: Estimator values of  $\theta_i^{(p)} = (\alpha_i^{(p)}, \beta_i^{(p)})$ , for all  $i \in \mathcal{L}$ , under  $KL_m$ ,  $H$  and  $\chi^2$  divergences.

	$KL_m$	$H$	$\chi^2$
Pearson coefficient	0.81	0.81	0.80

Tab. 7: Pearson correlation between  $\{\alpha_i^{(p)}, i \in \mathcal{L}\}$  and  $\{E_i, i \in \mathcal{L}\}$

	$KL_m$	$H$	$\chi^2$
MSE	0.00564	0.00565	0.00575

Tab. 8: Global MSE

## C Estimation of the conditional loading distribution given the unloading

In this Appendix we present very briefly the estimation method as it is similar to the one in Appendix B.

For any given unloading state  $j \in \mathcal{L}$ , assume we observe an i.i.d. sample  $(Y_\ell^{(j)}, \ell \in \mathcal{L} \setminus \{j\})$  with common and unknown unloading conditional probability distribution,  $P_{\theta_j^{(d)}}$ . Assume also that  $\theta_j^{(d)} \in \Theta_j^{(d)}$ , an open set of  $\mathbb{R}_+ \times \mathbb{R}_+$ . Define  $Q_i(j)$  the empirical unloading distribution given that loading pertaining to these samples given by Eq (12), namely

$$Q_i(j) = F_{i|j} = T_{ij}/T_{\bullet j}, i \in \mathcal{L} \setminus \{j\}.$$

For any unloading state  $j \in \mathcal{L}$ , the parametric statistical model considered is the following

$$\mathcal{M}_j^{(d)} = \left\{ P_{\theta_j^{(d)}}, \theta_j^{(d)} \in \Theta_j^{(d)} \right\}, \quad (36)$$

with,

$$\Theta_j^{(p)} = \left\{ \theta_j^{(d)} := (\alpha_j^{(d)}, \beta_j^{(d)}) \right\} \subseteq \mathbb{R}_+ \times \mathbb{R}_+, \quad (37)$$

$$P_{\theta_j^{(d)}}(i) = \frac{U_{ji}^{(d)}}{\sum_{k \in \mathcal{L}, k \neq j} U_{jk}^{(d)}} = \frac{E_i^{\alpha_j^{(d)}} / C_{ji}^{\beta_j^{(d)}}}{\sum_{k \in \mathcal{L}, k \neq j} E_k^{\alpha_j^{(d)}} / C_{jk}^{\beta_j^{(d)}}}, i \in \mathcal{L} \setminus \{j\} \quad (38)$$

The model (36)-(37)-(38) is identifiable, so that the minimum of the  $\phi$ -divergences estimator of  $\theta_j^{(d)}$ , denoted by,  $\hat{\theta}_{j,\varphi}^{(d)}$  is defined as the unique solution of the minimisation problem, for any given

unloading zone  $j \in \mathcal{L}$

$$\hat{\theta}_{j,\varphi}^{(d)} = \operatorname{argmin}_{\theta_j^{(d)} \in \Theta_j^{(d)}} \phi \left( P_{\theta_j^{(d)}}, Q_i(j) \right) = \operatorname{argmin}_{\theta_j^{(d)} \in \Theta_j^{(p)} \atop i \in \mathcal{L} \atop i \neq j} \sum \varphi \left( \frac{P_{\theta_j^{(d)}}(i)}{Q_i(j)(i)} \right) Q_i(j),$$

where,  $\phi \left( P_{\theta_j^{(p)}}, Q_i(j) \right) = \hat{\phi} \left( P_{\theta_j^{(d)}}, P_{\theta_i^{(d)}} \right)$ , with  $Q_i(j)$  given by Eq. (12).

## Estimation results

From Tabs. 9-10-11 the same conclusion is made as in Appendix B. As a result, with a theoretical argument of robustness and similar goodness of fit between the three criteria for  $\theta^{(p)}$  and  $\theta^{(d)}$ , the most suitable criterion is Hellinger's. For this reason, we have chosen this criterion for the numerical simulations of Section 7.

NUTS	Names	$KL_m$		$H$		$\chi^2$		$KL_m$	$H$	$\chi^2$
		$\alpha_j$	$\beta_j$	$\alpha_j$	$\beta_j$	$\alpha_j$	$\beta_j$			
BE10	Brussels	0.67	1.28	0.66	1.30	0.65	1.38	0.0285	0.0220	0.0250
BE21	Antwerp	1.28	1.23	1.30	1.30	1.35	1.51	0.0435	0.0372	0.0451
BE22	Limburg	1.08	2.21	1.15	2.29	1.34	2.61	0.0542	0.0475	0.0547
BE23	East Flanders	1.62	0.70	1.67	0.75	1.79	0.83	0.0429	0.0386	0.0514
BE24	Flemish Brabant	0.96	0.65	0.97	0.65	0.98	0.63	0.0474	0.0422	0.0556
BE25	West Flanders	1.55	1.24	1.67	1.23	1.85	1.24	0.0365	0.0339	0.0501
BE31	Walloon Brabant	0.16	1.36	0.19	1.43	0.27	1.60	0.0326	0.0266	0.0324
BE32	Hainaut	0.70	0.40	0.72	0.51	0.80	0.76	0.0489	0.0536	0.0593
BE33	Liège	0.68	1.94	0.77	1.94	1.02	2.09	0.0434	0.0383	0.0495
BE34	Luxembourg	1.00	0.10	1.00	0.10	1.00	0.10	0.0592	0.0483	0.0323
BE35	Namur	1.00	3.61	1.08	4.00	1.11	4.00	0.0240	0.0236	0.0439

Tab. 9: Estimator values of  $\theta_j^{(d)} = (\alpha_j^{(d)}, \beta_j^{(d)})$ , for all unloading zones  $j \in \mathcal{L}$ , under  $KL_m$ ,  $H$  and  $\chi^2$  divergences.

	$KL_m$	$H$	$\chi^2$
Pearson coefficient	0.73	0.72	0.72

Tab. 10: Pearson correlation between  $\{\alpha_j^{(d)}, j \in \mathcal{L}\}$  and  $\{E_j, i \in \mathcal{L}\}$

	$KL_m$	$H$	$\chi^2$
MSE	0.0419	0.0374	0.0454

Tab. 11: Global MSE

## D Stochastic transportation cost

The measure we use to evaluate the distance of the conditional Huff model we have proposed in Section 6.2.1 is the generalised transport cost between two centroids  $i$  and  $j$ , noted  $C_{ij}$  in the presence of uncertainties. We can refer to Combes et Lafourcade (2005) and reference therein for the bibliography on the subject in the non stochastic framework. The transport cost represents the expenditure required to transport goods between a loading location and a delivery location. This measure includes both

distance-related costs and time-related costs. For the costs related to the time duration, we consider two time units: the hour and the day. The definition we use is the one used by the French National Road Comity (CNR), the so-called "trinome's formula".

### Sources of uncertainties in transportation cost modeling

The sources of uncertainty are outlined in Tab. 12.

Variables	Distributions
Speed ( $V$ )	<b>Weibull</b> shape : $k = 36.57189$ scale : $\lambda = 65.99397$
Working hours per day ( $W$ )	<b>Translated Beta</b> shape1 : $k_1 = 5.039663$ shape2 : $k_2 = 2.918150$ lower Bound : $m = 9$ upper Bound : $M = 12$

Tab. 12: Transport cost uncertainties

### Transportation cost formulation

The transportation cost depends on reference costs related to distance between two centroids and time measured in hours and days. In 2018, the CNR updated its road trucking Belgian survey of companies dating back to 2013. This new study was carried out in 2017 among trucking companies, drivers and professional organisations. The vehicles chosen by the CNR to carry out long-term statistics on the operating costs of road carriers are 40-tonne articulated vehicles, considered representative of industrial goods transported by road. Tabs. 13-14-15 list annual averages and statistics reported by the CNR's Belgian trucking survey of the various components composing the cost.

**Distance reference cost.** The distance-related reference costs incurred when connecting the centroids of areas  $i$  and  $j$  is related to the distance  $d_{ij}$ , and the cost per kilometer  $C_k$ , as follows

$$d_{ij} C_k, \quad (39)$$

where  $C_k$  is expressed in terms of the yearly mileage of vehicle, fuel price, maintenance and toll costs as given in Tab. 13.

**Hour-time reference cost.** Similarly, we define the time-related costs per hour in terms of the random route travel time from  $i$  to  $j$ ,  $\Delta T_{ij}$ , and the cost per hour  $C_h$ , namely,

$$\Delta T_{ij} C_h, \quad (40)$$

with  $\Delta T_{ij} := d_{ij}/V$ , where  $V$  is the random speed given in Tab. 12, and the cost per kilometer  $C_h$  is expressed in terms of driving time per year, wages/primes, employer's contributions, and travel allowance as provided by Tab. 14.

**Remark.** For the estimation parameters procedure, we don't take into account the loading and the unloading times. If any, we should have instead a service time (route travel time plus loading and unloading times following the CNR's definition), such that  $d_{ij}/V + S_\ell + S_u$ , with  $S_\ell$  and  $S_u$  the loading and the unloading times.



**Daily-time reference cost.** The daily-time related reference costs is expressed in terms of route travel time  $\Delta T_{ij}$ , the random number of working hours per week  $W$  (Tab. 12), and the cost per day  $C_d$ , namely,

$$\frac{\Delta T_{ij}}{W} C_d, \quad (41)$$

where  $\Delta T_{ij}$  is defined as above, and  $C_d$  is the cost per day expressed in terms of working days per year, taxes, insurances, vehicle ownership and structure costs as provided by Tab.15.

**Total transport cost.** Finally, the total cost connecting the centroids of  $i$  and  $j$  is given by the sum of (39), (40), and (41), i.e.,

$$C_{ij} := d_{ij} C_k + \Delta T_{ij} C_h + \frac{\Delta T_{ij}}{W} C_d.$$

Since both the driving time  $\Delta T_{ij}$  and the working hours per day are non-negative random variables the total cost  $C_{ij}$  is a non-negative random variable.

<b>Table 1. Distance-related reference costs (2017)</b>	
	Average values
Annual mileage (km/year)	117 000 (1)
Fuel (€/year)	30 134 (2)
Maintenance (€/year)	11 115 (3)
Toll (€/year)	8000 (4)
<b>Cost per kilometre (<math>C_k</math> in €/km)</b> <b>((2)+(3)+(4))/(1)</b>	<b>0.42</b>

Source: CNR (2017)

Tab. 13: Average annual cost associated with the distance reference cost

<b>Table 2. Time-related reference costs per hour (2017)</b>	
	Average values
Driving time (h/year)	1672 (1)
Wages/Primes (€/year)	29 942 (2)
Employer's contribution	0.6056 (3)
Travel allowance (€/year)	8 211 (4)
<b>Cost per hour (<math>C_h</math> in €/h)</b> <b>((2)+(2)×(3)+(4))/(1)</b>	<b>33.66</b>

Source: CNR (2017)

Tab. 14: Average annual cost associated with the hourly reference cost

<b>Table 3. Time-related reference costs per day (2017)</b>	
	Average values
Number of working days (d/year)	220 (1)
Axle tax (€/year)	515 (2)
Insurance (vehicle) (€/year)	1900 (3)
Insurance (goods) (€/year)	250 (4)
Vehicle ownership (€/year)	14 769 (5)
Synthesis-cost(except structural costs)	126 658 (6)
Structure (%)	9.1 (7)
<b>Cost per day (<math>C_h</math> in €/d)</b> <b>((2)+(3)+(4)+(5)+((6)×100)/(100-(7)))/(1)</b>	<b>136.88</b>

Source: CNR (2017)

Tab. 15: Average annual cost associated with the daily reference cost

## References

- BAILEY, E., UNNIKRISHNAN, A. et LIN, D.-Y. (2011). Models for minimizing backhaul costs through freight collaboration. *Transportation Research Record*, 2224(1):51–60.
- BEN-AKIVA, M., SHERMAN, L. et KULLMAN, B. (1978). Disaggregate travel demand models for the san francisco bay area. non-home-based models. *Transportation Research Record*, (673).
- BORGERS, A. et TIMMERMAN, H. (1986). A model of pedestrian route choice and demand for retail facilities within inner-city shopping areas. *Geographical analysis*, 18(2):115–128.
- CHEN, R., DONG, J.-X. et LEE, C.-Y. (2016). Pricing and competition in a shipping market with waste shipments and empty container repositioning. *Transportation Research Part B: Methodological*, 85:32–55.
- COMBES, P.-P. et LAFOURCADE, M. (2005). Transport costs: measures, determinants, and regional policy implications for france. *Journal of economic geography*, 5(3):319–349.
- CONVERSE, P. D. (1949). New laws of retail gravitation. *Journal of Marketing*, 14(3):379–384.
- CRAINIC, T. G., GENDREAU, M. et DEJAX, P. (1993). Dynamic and stochastic models for the allocation of empty containers. *Operations research*, 41(1):102–126.
- DEJAX, P. J. et CRAINIC, T. G. (1987). Survey paper—a review of empty flows and fleet management models in freight transportation. *Transportation science*, 21(4):227–248.
- ERERA, A. L., MORALES, J. C. et SAVELSBERGH, M. (2009). Robust optimization for empty repositioning problems. *Operations Research*, 57(2):468–483.
- FAN, L., LUO, M. et WILSON, W. W. (2014). Pricing joint products in liner shipping. *International Journal of Shipping and Transport Logistics* 5, 6(4):371–386.
- FERRELL, W., ELLIS, K., KAMINSKY, P. et RAINWATER, C. (2020). Horizontal collaboration: opportunities for improved logistics planning. *International Journal of Production Research*, 58(14):4267–4284.
- GORMAN, M. F. (2001). Intermodal pricing model creates a network pricing perspective at bnsf. *Interfaces*, 31(4):37–49.
- GORMAN, M. F. (2002). Pricing and product mix optimization in freight transportation. *Transportation Quarterly*, 56(1).

- HOLMBERG, K., JOBORN, M. et LUNDGREN, J. T. (1998). Improved empty freight car distribution. *Transportation Science*, 32(2):163–173.
- HUFF, D. L. (1964). Defining and estimating a trading area. *Journal of Marketing*, 28(3):34–38.
- JIMÉNZ, R. et SHAO, Y. (2001). On robustness and efficiency of minimum divergence estimators. *Test*, 10(2):241–248.
- JOBORN, M., CRAINIC, T. G., GENDREAU, M., HOLMBERG, K. et LUNDGREN, J. T. (2004). Economies of scale in empty freight car distribution in scheduled railways. *Transportation Science*, 38(2):121–134.
- LERMAN, S. R. (1979). The use of disaggregate choice models in semi-markov process models of trip chaining behavior. *Transportation Science*, 13(4):273–291.
- LI, J.-A., LEUNG, S. C., WU, Y. et LIU, K. (2007). Allocation of empty containers between multi-ports. *European Journal of Operational Research*, 182(1):400–412.
- LI, J.-A., LIU, K., LEUNG, S. C. et LAI, K. K. (2004). Empty container management in a port with long-run average criterion. *Mathematical and Computer Modelling*, 40(1-2):85–100.
- LIESE, F. et VAJDA, I. (1987). *Convex statistical distances*, volume 95. Teubner.
- LINDSAY, B. G. (1994). Efficiency versus robustness: the case for minimum hellinger distance and related methods. *The annals of statistics*, 22(2):1081–1114.
- LONG, Y., LEE, L. H. et CHEW, E. P. (2012). The sample average approximation method for empty container repositioning with uncertainties. *European Journal of Operational Research*, 222(1):65–75.
- LU, T., LEE, C.-Y. et LEE, L.-H. (2020). Coordinating pricing and empty container repositioning in two-depot shipping systems. *Transportation Science*, 54(6):1697–1713.
- LUCE, R. D. (1959). *Individual Choice Behavior: A Theoretical analysis*. Wiley, New York, NY, USA.
- MORALES, D., PARDO, L. et VAJDA, I. (1995). Asymptotic divergence of estimates of discrete distributions. *Journal of statistical Planning and Inference*, 48(3):347–369.
- ÖZENER, O. Ö. et ERGUN, Ö. (2008). Allocating costs in a collaborative transportation procurement network. *Transp. Sci.*, 42:146–165.
- PAN, S., TRENTESAUX, D., BALLOT, E. et HUANG, G. Q. (2019). Horizontal collaborative transport: survey of solutions and practical implementation issues. *International Journal of Production Research*, 57(15-16):5340–5361.
- PEDERSON, L. E., HAMMER, R. M. et CASAVANT, K. L. (1979). Factors affecting interstate backhauling of exempt agricultural commodities by regulated motor carriers: a first look. *Transportation Journal*, pages 46–52.
- PIGOU, A. et TAUSSIG, F. (1913). Railway rates and joint costs. *The Quarterly Journal of Economics*, 27(4):687–694.
- REILLY, W. (1931). The law of retail gravitation.
- REILLY, W. J. et al. (1929). Methods for the study of retail relationships.
- SONG, D.-P. (2007). Characterizing optimal empty container reposition policy in periodic-review shuttle service systems. *Journal of the Operational Research Society*, 58(1):122–133.
- SONG, D.-P. et CARTER, J. (2009). Empty container repositioning in liner shipping. *Maritime Policy & Management*, 36(4):291–307.

- THILL, J.-C. et THOMAS, I. (1987). Toward conceptualizing trip-chaining behavior: A review. *Geographical Analysis*, 19(1):1–17.
- THILL, J.-C. et TIMMERMANS, H. (1992). Analyse des décisions spatiales et du processus de choix des consommateurs: théorie, méthodes et exemples d'applications. *L'Espace géographique*, pages 143–166.
- TOMA, A. et BRONIATOWSKI, M. (2011). Dual divergence estimators and tests: robustness results. *Journal of Multivariate Analysis*, 102(1):20–36.
- XU, L., GOVINDAN, K., BU, X. et YIN, Y. (2015). Pricing and balancing of the sea–cargo service chain with empty equipment repositioning. *Computers & Operations Research*, 54:286–294.
- ZHOU, W.-H. et LEE, C.-Y. (2009). Pricing and competition in a transportation market with empty equipment repositioning. *Transportation Research Part B: Methodological*, 43(6):677–691.
- Özlem ERGUN, KUYZU, G. et SAVELSBERGH, M. (2007). Shipper collaboration. *Computers & Operations Research*, 34:1551–1560.