Junior Management Science 9(1) (2024) 1241-1268

Junior Management Science

www.jums.academy
ISSN: 2942-1861

Cost Allocation in Vehicle Routing Problems with Time Windows

Federico Arroyo

Technical University of Munich

Abstract

The estimation of costs allocated to each customer when serving them in a collaborative logistic operation is a complex problem whose solution is computationally very expensive. In this work the case of central horizontal collaboration for vehicle routing problems with time windows and a central depot is studied. An approximation to the Shapley value method via structured random sampling is used to calculate the cost associated with customers in Solomon instances. Such costs are regressed to a linear model with a set of defined features. The results show that cost can be predicted with considerable accuracy with few features. Moreover, the extent to which vehicles' capacity, customers' demand and distance, the degree of customer clustering and time window horizons affect cost and potential savings from carriers in collaboration is assessed. Additionally, individual regression models of different set of instances show how various pricing strategies for customers can be fitted to their classification when grouping them.

Keywords: collaborative vehicle routing; cost allocation; Shapley value method; structured random sampling; time windows

1. Introduction

1.1. Context and motivation

With the world's population continuing to grow and an ever-expanding international commerce, logistics networks become increasingly complex. The supply chain, as a means of delivering products from manufacturers to customers, comprises the interactions between enterprises involving a multi-organizational network adding value to the flow of information, logistics and capital (Leng et al., 2018; Tang et al., 2022). It is within this operation that any potential for improvement brings substantial value.

Further, the mass adoption of the internet, in conjunction with the development of new technologies, has made online shopping a very comfortable means for consumers to acquire their products, favoring it in many cases over brick-and-mortar shops. This change of behavior towards ordering products online has meant rapid growth for e-commerce which has been further exacerbated by the COVID-19 pandemic, feeding the need for more retailers online and the growth in deliveries (Alfonso et al., 2021; Jílková & Králová, 2021). The final leg of the logistics problem, known as the last mile, is where retailers and logistic carriers mostly find their costs increasing due to the inefficiencies of the industry (Iannaccone et al., 2021). This operation of delivering a parcel from the last hub to the customer accounts for 28 % of the total logistic cost (Ranieri et al., 2018). Moreover, the effect of increased urban parcel demands in the context of urbanization can be very harmful to cities. With 70 % of the world's population expected to live in cities by 2050, the impact on the environment, safety and health from pollution and congestion becomes even more important (Bretzke, 2013). Consequently, public institutions and governments are beginning to regulate carriers' operation more intensely, pushing for in-

I would like to express my gratitude to Prof. Dr. Stefan Minner for affording me the invaluable opportunity to compose my master thesis under his supervision and mentorship. His continuous provision of insightful suggestions, innovative ideas to enhance my thesis, and unwavering support during the writing process have been truly instrumental. Furthermore, my appreciation extends to the esteemed Chair of Logistics and Supply Chain Management of the Technical University of Munich where I was given full assistance and support along my studies and in preparation for my thesis. Lastly, I extend my thanks to family and friends, whose steadfast encouragement and caring have been a driving force behind my endeavors, fostering both my personal and professional growth.

DOI: https://doi.org/10.5282/jums/v9i1pp1241-1268

creased sustainability and efficiency in their operations (Hu et al., 2019). With regards to the timing of deliveries, first-time delivery failures have been reported up to 60 %, resulting in a significant cost for carriers (Song et al., 2009). Additionally, same-day deliveries and the expectations coming from customers of receiving their parcels fitting their schedules compel agents to work with tight time horizons.

Within this context, companies are either forced or need to collaborate with each other to remain competitive and streamline their costs. Horizontal collaboration appears as an effective solution for carriers to increase their operational efficiency (Gansterer & Hartl, 2018b). For this reason, it is increasingly important for public agents in the last years (Cruijssen & Cruijssen, 2020). In a sharing economy, where companies are using common resources and capacities in urban areas to provide new services, new business models emerge and traditional ones need to innovate in order to continue doing business (Dahle et al., 2019).

1.2. Problem statement

In recent years, the area of collaborative vehicle routing has gained great popularity in transportation and logistics research (Gansterer & Hartl, 2020). Shared transportation resources define the base concept where different partners join in a common operation. In the realm of collaborative vehicle routing, we focus on those problems with capacity and time constraints, later explained in Section 3.1. To cope with the growth of deliveries and the ever-tightening constraints of time and capacity, collaborative environments allow for a great level of optimization within logistic operations (Karaenke et al., 2019; Vanovermeire et al., 2014). In this regard, there is a strong incentive for carriers to participate in a joint coalition given the cost reductions they can achieve (Muñoz-Villamizar et al., 2015).

Moreover, through the optimized use of resources, companies can benefit from providing better service, due to their increased flexibility and better times of delivery (Cruijssen, Dullaert, & Fleuren, 2007). Additionally, there is a need for them to decrease their environmental impact. In the USA, for example, it has been found that logistics vehicles can travel totally empty in up to 15% to 20% of the distance they cover (Ferrell et al., 2020). Companies, therefore have the motivation to increase their sustainability and their reputation.

Within the paradigm of companies cooperating and sharing resources, we focus on horizontal collaboration where carriers belong to the same level of the supply chain. Vertical collaboration, on the other hand, involves other levels in it. A common understanding, represented through a central system or platform, is needed for agents to interact and plan together horizontally, with some cities already introducing such systems (Schmelzer et al., 2016). The way collaboration structures are defined is crucial to allow for horizontal collaborations to prosper due to competitors being reticent to share their information and resources for lack of trust (Bratton et al., 2000). Transparency and confidentiality are key for competitors not to fear their strategic advantage is threatened. Consequently, having a fair distribution of costs arisen and profits obtained is of greater importance for the stakeholders involved (Cruijssen, Cools, & Dullaert, 2007). In this regard, our work is centered in cost allocation in collaborative vehicle routing problems, where the total cost of a global operation needs to be divided between the individual carriers. The objective is to assign costs to each of the customers that each partner brings into the coalition in a simple but fair way. The cost allocation method must assure that no partner perceives their allocated cost as being too large, for they risk the stability of the cooperation. We explore this in Section 3.2. The numerical complexity of the problem calls for the search for simpler mechanisms which is at the center of our work.

1.3. Research question and thesis outline

Given the scope presented, we aim at answering the following question:

> *Can cost allocation mechanisms in vehicle routing problems with time windows be approximated through simple methods?*

With this, we determine the accuracy of a simple cost-allocating linear model where certain features are used to address the variables of the problem. Additionally, by using different instances with different degrees of clustering and time window horizons, we analyze the impact these dimensions have on the cost-allocating process and derive conclusions for different pricing strategies. Concerning the allocation mechanism itself, we determine the degree to which the allocating results are advantageous for all carriers and whether or not all customers yield savings when shared with others.

This work begins by reviewing the available literature on the subject and summarizing basic definitions in Section 2. The methodology to tackle the problem in question is defined in Section 3 where the modelling and solving of vehicle routing problems, cost-allocating mechanisms and regression methods are presented. Later, in Section 4, the analyzed instances and cost function features are introduced. Subsequently, the results obtained are analyzed in Section 5. This comprises an overview of the costs allocated to each customer for the different instances, followed by the model fit and feature selection and the study of collaborative gains. To finalize, conclusions and recommendations for future research directions are pointed out in Section 6.

2. Literature review

In the context of our investigation, we begin our work by thoroughly reviewing the current state of academic research concerning cost allocation in collaborative multi-carrier vehicle routing problems with time windows. To better digest the subject, we focus on the two main bodies of research that our study focuses on. On the one hand, we start by evaluating the vehicle routing problem and its variations; focusing on horizontal collaboration. Further, we describe the related work in the area of cost allocation methods for collaborating logistics. We aim to present the known sources on which we base our work, finding potential unexplored areas in research that we intend to develop.

2.1. The vehicle routing problem

The Vehicle Routing Problem (VRP), introduced by Dantzig and Ramser (1959) and Clarke and Wright (1964), still remains one of the most widely studied problems in operations research (Kritikos & Ioannou, 2010). It consists of a family of problems in which a set of customers with known demands and geographical location have to be served by a fleet of vehicles based at one (or many) depot(s) through a network of paths between them. The task is to determine the optimal set of routes for the vehicles which achieve the objective and satisfy the constraints imposed on the problem (Irnich et al., 2014).

Multiple variations and extensions to the VRP have been introduced which are mainly classified based on their objectives and constraints (Laporte, 2009). Most commonly, a capacity restriction is imposed on the vehicles, with each location having a specified demand quantity. This problem, known as the Capacitated VRP (CVRP), imposes that each customer needs to be served only once and the capacity of vehicles is not surpassed by the total amount of customer demands served by the vehicle in its route (Goel & Gruhn, 2008; Prins, 2004). Additionally, the CVRP with Time Windows (CVRPTW) introduces the time dimension and the constraint of customers being served in a time interval, also known as Time Window (Kallehauge et al., 2005). The addition of several depots with multiple pickup and delivery points (Berbeglia et al., 2007) as the discrimination between full truckload and less than truckload (Gansterer & Hartl, 2018b) are also well-known alternatives that we will not explore in our work.

When several carriers are present, each of them having to solve a VRPs by their own, potential benefits have been found when collaborating in a joint logistic operation, originally reviewed in (Cruijssen, Dullaert, & Fleuren, 2007; Krajewska & Kopfer, 2006).

In their survey, Gansterer and Hartl (2018b) show that problems in vehicle routing in a collaborative scenario can be addressed from a central or decentral planning perspective. Central plans require the intervention of one decision-maker who is fully informed. The information provided for each carrier can consist of cost structures, capacities and information on existing customers (Gansterer & Hartl, 2018b). An example of this in digitalized times would be an online platform/database where carriers' requests are shared and solved in a logistics problem. In this situation, the problem is reduced to a holistic standard optimization problem where all resources are bundled and the collaborative aspect is overtaken by the full disclosure of information.

In decentralized collaborations, carriers are able to share their capacities to a certain extent by revealing a limited amount of information. In said setting partners may deal individually with each other or channel their requests through a central authority. Studies on this area are presented by (Berger & Bierwirth, 2010; Dai & Chen, 2011; Gansterer & Hartl, 2018a; Gansterer et al., 2019; Krajewska & Kopfer, 2006; Özener et al., 2011), covering non-auction and auction-based mechanisms for the selection and exchange of requests. As reviewed by Gansterer and Hartl (2018b), when it comes to real-world cases the potential collaboration gains have been calculated up to 20%-30% (Chinh et al., 2016; Cruijssen, Bräysy, et al., 2007; Ergun et al., 2007; Muñoz-Villamizar et al., 2015). In the case of each customer requiring service by two or more carriers Fernández et al. (2018) obtained cost savings of $6\% - 25\%$. Next to the economical aspect, the nature of collaboration between carriers signifies a reduced negative impact on the environment. These aspects are found in the research by Ballot and Fontane (2010), Muñoz-Villamizar et al. (2015), Pérez-Bernabeu et al. (2015), and Schulte et al. (2017) which study the reduction of routes and vehicles leading to a decrease in CO2 emissions and congestion in cities.

From the numerical solving perspective, collaborative VRPs encompass a great complexity which requires the use of heuristic and metaheuristic approaches. With regard to the VRP, the solving mechanisms are found in Cordeau et al. (1997, 2007), Laporte (1992), and Tarantilis (2005). Collaborative VRPs solving algorithms are discussed by Defryn, Sörensen, and Cornelissens (2016), Pérez-Bernabeu et al. (2015), and Sanchez et al. (2016).

2.2. Cost allocation methods

In the scenario of collaboration, the profits gained by the coalition and the costs assumed have to be shared among the participants. A transparent and fair mechanism to split these values is necessary in order to assure the collaborative relationships between the partners and their long-term participation (Defryn, Sörensen, & Cornelissens, 2016).

In the context of the VRP and its variations, the cost allocation problem consists of splitting the cost of the route/s between the customers served. This was first addressed by Göthe-Lundgren et al. (1996). Some other practical applications were developed by Engevall et al. (2004), Krajewska and Kopfer (2006), and Krajewska et al. (2008). For CVRPTW, cost allocation was first studied by Cruijssen et al. (2010) and later analyzed by Dahl and Derigs (2011) in a dynamic setting.

In their survey, Guajardo and Rönnqvist (2016) show that in most cases in literature cooperative game theory is applied for the sharing of costs or profits. The authors categorize more than 40 methods either as traditional or ad-hoc. The first one refers to methods arising from previous work on cooperative game theory whereas ad-hoc are methods which result out of the particular conditions of the problem. Nonetheless, they conclude that a big majority of the methods found in literature fall into the following:

- the Shapley value (Shapley, 1953), generally the most applied method which we explain later;

- proportional methods, where the cost is divided proportionally to a variable to each carrier (e.g. carrier j is allocated cost α_j). The share of cost assigned to each player can follow different criteria (e.g. shared equally among players, according to their demand quantities or standalone costs);
- the nucleolus method (Schmeidler, 1969), which looks for an individually rational distribution in which the maximum dissatisfaction is minimized.

Table 1 extracted from the literature survey by Guajardo and Rönnqvist (2016) summarizes the methods that are of most relevance for our work. From the traditional methods, the Shapley value is the one we focus on given its broad use. The methods consolidated under the category "ad-hoc" respond to those following definitions which concern the particular context where the research has been motivated, including modifications of traditional methods.

In their work Shi et al. (2020) analyse the problem of sharing profits in collaborative vehicle routing problems with multiple depots. A similar approach is what we apply for splitting costs in problems with rather one depot and the added constraint of time-windows. Akkerman and Mes (2022) analyze how customer selection in vehicle routing problems could be approximated by the distance. We expand this area of study by addressing which other features, and to which extent, could be used to approximate cost being allocated to customers. To our best knowledge, there is no academic work where the cost allocation for collaborative environments of CVRPTW is addressed through approximation methods using linear regression. This is what we wish to explore with our work.

3. Methodology

As the focus of this work is to analyze the applicability of direct rules and approximate a formula for the cost allocation of clients in a collaborative vehicle routing problem, we will develop our work by evaluating different CVRPTW Solomon data instances (Solomon, 1987) and assessing how the measures designed fit the outcome obtained. Firstly, we model the logistics problem as a vehicle routing problem with the added constraints of vehicle capacity and time windows and build our optimization model. In order to streamline the computation of our results, a metaheuristic solver is introduced. Further, we assess different methods for the allocation of costs in collaborative games. In particular, the full calculation of the Shapley value is initially adopted and later replaced with a structured random sample-based method. Finally, our model is applied to existing problem instances so as to derive the results which are later analyzed.

3.1. Modelling and solving of vehicle routing problems

3.1.1. Capacitated Vehicle Routing Problem (CVRP)

For the modelling of our problem, we start with a simpler case in the CVRP. In this regard, the two-index flow formulation by Toth and Vigo (2014) is used. The CVRP aims at finding the set of routes that allow all customers' demands in a network to be serviced by a fleet of vehicles starting from a depot with the minimum cost. Further, customers' demands and geographical locations are deterministic with the latter deriving in the network of possible arcs available between nodes. Vehicles have a limit to the capacity that they can carry. In this regard, the notation is defined in the following.

Let $V = 0, \ldots, n$ be the vertex set, where vertices $i = 1, \ldots, n$ correspond to the customers and vertex 0 corresponds to the depot. A is then the arc set which completes the graph $G = (V,A)$, with each arc having a cost c_{ij} associated with it: $(i, j) \in A$. Furthermore, each customer has a demand d_i to be fulfilled by a set of K available vehicles (K being not smaller than the minimum amount of vehicles needed to serve all customers), which are indistinguishable with the same capacity C measured in the same units as the demand. For a given circuit, the cumulative demand of all the customers served by it must not exceed the vehicle capacity. Finally, given a subset $S \subseteq V\backslash\{0\}$, we denote by $r(S)$ the minimum number of vehicles needed to serve all customers in S.

In all the instances we use throughout our work, vertices are defined by points with specific coordinates. Further, the cost c_{ij} linked to each arc $(i, j) \in A$, is calculated as the Euclidean distance between the two points corresponding to the vertices i and j. This results in a cost matrix which is symmetric and satisfies the triangle inequality:

$$c_{ik} + c_{kj} \geq c_{ij} \quad \forall i, j, k \in V$$

which signifies that any deviation from the direct link between two vertices results in a bigger distance, therefore bigger cost. In the remaining of this work, distance will be used as a measure of cost and vice-versa.

The model uses binary variables x_{ij} which take the value of 1 if the arc (i, j) is traversed in the solution and 0 otherwise. With the notation described, the problem formulation is as follows:

$$\min \sum_{i\in V}\sum_{j\in V} c_{ij}x_{ij} \tag{1a}$$

$$\text{s.t.:} \sum_{i\in V} x_{ij} = 1 \quad \forall j \in V\backslash\{0\}, \tag{1b}$$

$$\sum_{j\in V} x_{ij} = 1 \quad \forall i \in V\backslash\{0\}, \tag{1c}$$

$$\sum_{i\in V} x_{i0} = K, \tag{1d}$$

$$\sum_{j\in V} x_{0j} = K, \tag{1e}$$

$$\sum_{i\notin S}\sum_{j\in S} x_{ij} \geq r(S) \quad \forall S \subseteq V\backslash\{0\}, S \neq \emptyset, \tag{1f}$$

$$x_{ij} \in \{0, 1\} \quad \forall i, j \in V. \tag{1g}$$

The objective function in equation 1a aims at minimizing the sum of costs deriving from traversing the routes. Constraints 1b and 1c ensure that each customer is visited only

Table 1: Cost allocation methods in a collaborative environment found in academic literature (Guajardo & Rönnqvist, 2016).

Method	No. of articles	References
Ad hoc	31	(Agarwal & Ergun, 2010; Anily & Haviv, 2007; Audy et al., 2011, 2012; Caprara & Letchford, 2010; Cheng et al., 2013; Dahl & Derigs, 2011; Dai & Chen, 2012, 2015; Derks & Kuipers, 1997; Engevall et al., 1998, 2004; Estévez-Fernández et al., 2009; Faigle et al., 1998; Fiestras-Janeiro et al., 2013; Fiestras-Janeiro et al., 2012; Flisberg et al., 2015; Frisk et al., 2010; Guajardo & Rönnqvist, 2015; Hamers et al., 1999; Hezarkhani et al., 2016; Liu et al., 2010; Özener, 2014; Özener & Ergun, 2008; Özener et al., 2013; Potters et al., 1992; Sun et al., 2015; Toriello & Uhan, 2013; Vanovermeire & Sörensen, 2014b; Yang et al., 2016; Yilmaz & Savasaneril, 2012)
Shapley	23	(Agarwal & Ergun, 2010; Cheng et al., 2013; Cruijssen et al., 2010; Dror, 1990; Engevall et al., 1998, 2004; Fang & Cho, 2014; Fiestras-Janeiro et al., 2012; Frisk et al., 2010; Guajardo et al., 2016; Hezarkhani et al., 2016; Krajewska et al., 2008; Liu et al., 2010; Lozano et al., 2013; Massol & Tchung-Ming, 2010; Özener et al., 2013; Sun et al., 2015; Vanovermeire & Sörensen, 2014a; Vanovermeire et al., 2014; Wang et al., 2015; Wong et al., 2007; Yengin, 2012; Zakharov & Shchegryaev, 2015)
Proportional	18	(Audy et al., 2012; Berger & Bierwirth, 2010; Dror, 1990; Engevall et al., 2004; Fishburn & Pollak, 1983; Flisberg et al., 2015; Frisk et al., 2010; Hezarkhani et al., 2016; Krajewska & Kopfer, 2006; Lehoux et al., 2011; Liu et al., 2010; Massol & Tchung-Ming, 2010; Nguyen et al., 2014; Özener, 2014; Özener & Ergun, 2008; Özener et al., 2013; Sun et al., 2015; Wong et al., 2007)

once allowing for exactly one arc to enter and leave each vertex associated with them. Consequently, constraints 1d and 1e impose analog requirements to the depot. Note that with this notation the CVRP aims at finding a collection of exactly K circuits to serve all customers, where each circuit starts and ends at the depot. Constraints 1f ensure connectivity of the paths in the solution while also imposing the vehicle capacity requirements. The constraints remain valid also if $r(S)$ is replaced by a trivial lower bound defined by $\lceil d(S)/C \rceil$ (Cornuejols & Harche, 1993) with $d(S) = \sum_{i \in S} d_i$ denoting the total demand of the set $S \subseteq V$. Lastly, constraint 1g represents the binary constraint.

The family of constraints 1f has a cardinality which grows exponentially with n (Toth & Vigo, 2014). A family of equivalent constraints with polynomial cardinality can be used by utilizing the sub-tour elimination constraint of the Travelling Salesman Problem (TSP) applied to the CVRP (Christofides, 1979; Desrochers & Laporte, 1991; Miller et al., 1960):

$$u_i - u_j + Cx_{ij} \leq C - d_j \quad \forall i, j \in V \backslash \{0\}, i \neq j \tag{2a}$$

$$\text{such that } d_i + d_j \leq C,$$

$$d_i \leq u_i \leq C \forall i \in V \backslash \{0\}, \tag{2b}$$

where the variable $u_i, i \in V \backslash \{0\}$ is added to the model, accounting for the load of the vehicle after visiting customer i. With this definition, when $x_{ij} = 0$, constraint 2a is not binding, while when $x_{ij} = 1$ then $u_j \geq u_i + d_j$ imposing the capacity and connectivity requirements and eliminating isolated sub-tours in the process (Toth & Vigo, 2014).

3.1.2. CVRP with Time Windows (CVRPTW)

The CVRP with Time Windows extends the CVRP further restricting the problem by adding a time dimension. Each customer is associated with a time interval in which it needs to be served, called a time window. Further, the travel time, t_{ij}, for each arc $(i, j) \in A$ is known and normally taken as the same as the distance or cost of the arc c_{ij}. Additionally, service times s_i are given for each customer i and the time instant in which the vehicles leave the depot is also known and normally assumed to be at 0. As a result, the service at each customer i has to be initiated during their time window and it lasts the service time associated with it s_i. Furthermore, whenever a vehicle arrives earlier than the start of a customer's time window, it is allowed to wait until it begins and then commence service.

For the purpose of this work and to summarize it's main characteristics, the classical CVRPTW consists of defining exactly one circuit per vehicle (of the total K vehicles available) minimizing the cost with:

- each vehicle starting and finishing at the depot,
- customers being serviced exactly once,
- the capacity of the vehicles C not exceeding the sum of the demands of the customers serviced by it,
- service starting within the time window $[a_i, b_i]$ of each customer and the vehicle remaining a time equal to s_i at location i.

The model used in our work for the CVRPTW follows the formulation by Toth and Vigo (2014), which is here described. It is based on the three-index vehicle flow model which allows for the modelling of more constrained versions of the vehicle routing problem, due to being more flexible in integrating additional dimensions.

Firstly, a network $G = (V,A)$ is defined by the set of vertices of customers and the depot, which in this case is represented by the two nodes 0 and $n+1$. The set of customers N is then defined as all vertices but vertices 0 and $n+1$, namely $N = V\backslash\{0, n+1\}$. Further, for a route to be feasible then it must start from node 0 and end at node $n+1$. Additionally, a time window $[a_i, b_i]$ is associated with each customer i. Nodes 0 and $n+1$ are assigned E and L as their earliest possible departure and the latest possible arrival at the depot respectively; that is, $[a_0, b_0] = [a_{n+1}, b_{n+1}] = [E, L]$. Moreover, and in accordance with the problem description, zero demands and service times are defined for these two nodes, namely $d_0 = d_{n+1} = s_0 = s_{n+1} = 0$.

Let $\Delta^+(i)$ denote the set of vertices j that are connected to the vertex i directly, such that arc $(i, j) \in A$. In the same manner, $\Delta^-(i)$ denote the vertices from which i is directly reached, e.g. the set of vertices j such that arc $(j, i) \in A$.

For each arc $(i, j) \in A$ and vehicle $k \in K$, a variable x_{ijk} is introduced which takes the value of 1 if arc (i, j) is used by vehicle k and 0 otherwise. Further, a time variable w_{ik} is added for all nodes $i \in V$ and vehicles $k \in K$ which designates the time of the start of service at node i by vehicle k.

With the described notation, the formulation as of Toth and Vigo (2014) is as follows:

$$\min \sum_{k\in K} \sum_{(i,j)\in A} c_{ij} x_{ijk} \tag{3a}$$

$$\text{s.t.:} \sum_{k\in K} \sum_{j\in\Delta^+(i)} x_{ijk} = 1 \quad \forall i \in N, \tag{3b}$$

$$\sum_{j\in\Delta^+(0)} x_{0jk} = 1 \quad \forall k \in K, \tag{3c}$$

$$\sum_{i\in\Delta^-(j)} x_{ijk} - \sum_{i\in\Delta^+(j)} x_{jik} = 0 \quad \forall k \in K, j \in N, \tag{3d}$$

$$\sum_{i\in\Delta^-(n+1)} x_{i,n+1,k} = 1 \quad \forall k \in K, \tag{3e}$$

$$x_{ijk}\left(w_{ik} + s_i + t_{ij} - w_{jk}\right) \leq 0 \quad \forall k \in K, (i, j) \in A, \tag{3f}$$

$$a_i \sum_{j\in\Delta^+(i)} x_{ijk} \leq w_{ik} \leq b_i \sum_{j\in\Delta^+(i)} x_{ijk} \quad \forall k \in K, i \in N, \tag{3g}$$

$$E \leq w_{ik} \leq L \quad \forall k \in K, i \in \{0, n+1\}, \tag{3h}$$

$$\sum_{i\in N} d_i \sum_{j\in\Delta^+(i)} x_{ijk} \leq C \quad \forall k \in K, \tag{3i}$$

$$x_{ijk} \in \{0, 1\} \quad \forall k \in K, (i, j) \in A. \tag{3j}$$

The objective function calculates the total cost incurred which we aim to minimize. Constraints 3b ensure that each customer is assigned to exactly one route. Further, constraints 3c to 3e designate the flow on the path to be followed by vehicle k. After, constraints 3f to 3h ensure the time schedule is feasible while 3i guarantees the capacity constraint. Finally, constraints 3j ensure the binary conditions of the variables x_{ijk}.

3.1.3. Metaheuristic solver

The CVRP, as described in Section 3.1.1, is a generalization of the TSP (this being a CVRP with one vehicle visiting all vertices in a cost-minimizing circuit with $C \geq d(V)$ and $K = 1$). This problem and its extensions with Time Windows (CVRPTW) is known to be strongly NP-hard (Toth & Vigo, 2014). For this reason, it would be very expensive computationally to solve each problem to optimality. As a consequence, a solver algorithm using metaheuristics is implemented.

VROOM is an open-source optimization engine specifically tailored for VRPs which provides very good solutions in little time (VROOM Project, 2022). As an example of it performance, Table 2 shows the results obtained when solving Solomon 56 instances of 100 customers which are used for benchmarking. By comparing the cumulated traveled time, it is observed that the maximum gap difference with the best-known solution is 3.89%, all with an average computing time of 359 milliseconds.

The mentioned algorithm utilizes different heuristics depending on the problem to find an initial solution. In the case of the CVRPTW, it uses modified versions of Solomon insertion heuristics. To improve the solution, a local search procedure consisting of 14 different operators is then performed to check for better feasible neighboring solutions (Bräysy & Gendreau, 2005). In their work comparing tools for solving CVRPTW, Puka et al. (2019) find that VROOM in comparison with other open-source solvers is one of the most balanced in terms of solution quality and execution time. For all the aforementioned reasons, it is used in our research.

3.2. Cost allocation in VRPs in a collaborative environment

In this section, the vehicle routing problems previously defined are introduced in a collaborative environment in which several companies, each with a set of customers to serve, form a coalition that aims at satisfying the demand of all customers in a joint operation. In this setup, all customers and their demands are joined into a holistic logistic problem that is indifferent to the source of each customer. Operating as a joint venture, it has been shown that a total reduced cost is achieved while serving all customers. This is produced by the increased efficiency in the joint logistic operation due to customers with similar locations or service time windows being visited by the same vehicle. The cost that companies pay for such an optimization coalition is having to share their customer information which may be of high value for their business. We will consider a set of partners which form a grand coalition sharing their resources (vehicles) and demands. An example of such a collaborative CVRP is displayed in Figure 1 where 3 partners join in a coalition to serve all of their customers resulting in a completely different total logistic operation.

When working as a coalition of two or more partners, the problem arises of distributing the costs (or the profits) of serving the customers as a whole. In this regard, a mechanism is needed to allow for each customer that is brought

Table 2: Comparison to best known solutions (BKS) targeting cumulated travel time (CTT). Cumulated number of vehicles (CNV) is only reported for the record. (VROOM Project, 2022)

Class	C1	C2	R1	R2	RC1	RC2	Total
BKS CNV	10.0	3.0	13.25	5.36	12.88	6.25	485
BKS CTT	828.38	589.86	1175.75	878.41	1340.02	1004.21	54699
VROOM CNV	10.0	3.0	13.25	4.27	13.25	5.25	468
VROOM CTT	828.38	589.86	1191.99	912.60	1356.76	1030.75	55,616
CTT gap	+0.00%	+0.00%	+1.38%	+3.89%	+1.25%	+2.64%	+1.68

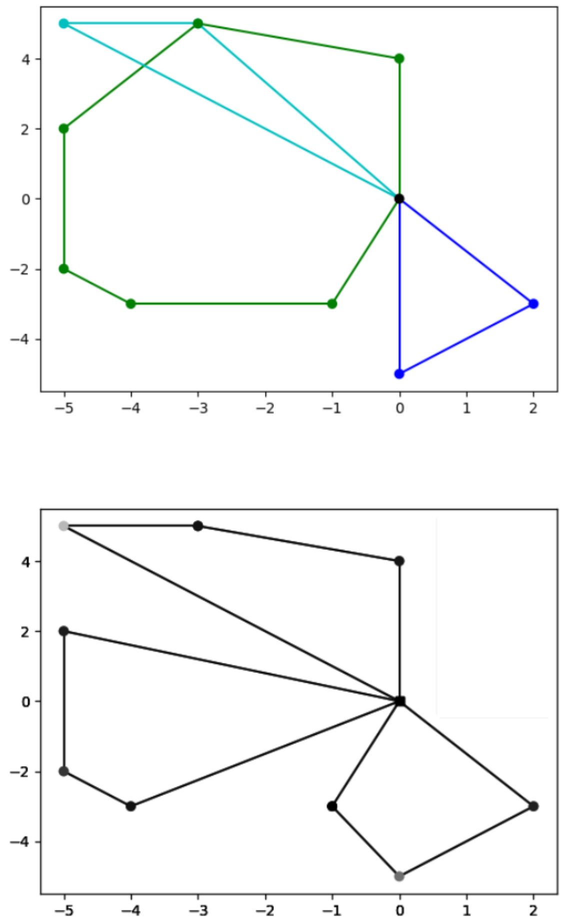


Figure 1: Example of a collaborative VRP where 3 partners (top green, cyan, blue) work jointly on a logistic operation.

in the coalition by a partner to be correctly accounted for the marginal cost it produced to the total operation. Despite no method being considered as a global best practice, Defryn, Vanovermeire, and Sörensen (2016) argue that a cost allocation method based on the incentives of each partner should be selected by the coalition. This results in the group behaving in the best interest of the whole, e.g.: a cost allocation based on volume will derive in benefit for those who transport the highest volumes therefore steering the coalition into increasing the total volume displaced. Nonetheless, there is a general consensus on the fact that incentive criteria results in a fair distribution of costs in all cases (Defryn, Sörensen, & Cornelissens, 2016). We follow with the introduction of the concepts behind cooperative game theory and an axiom-based cost allocation method: the Shapley value.

3.2.1. Cooperative game theory: Core

In game theory when evaluating a game, one is interested in knowing the value each player adds to a coalition when a group of players cooperates and obtains a certain overall gain from that cooperation. Some players' contributions may be larger or they may have a bigger bargaining power, even so to the extent of threatening to destroy the entire surplus. Therefore, the goal of this assessment would be to determine how important is each participant to the overall coalition, and what would therefore be their expected fair payoff. An analogous formulation applies to the case of determining the cost of each customer added to a collaborative logistic problem as the only adjustment is that in the VRP each added player (customer) to the coalition will add a cost for the total operation instead of a surplus.

Let us introduce some cooperative game theory concepts as described by Guajardo and Rönnqvist (2015). For a cooperative game in which the "grand coalition" of all players in the set $N = \{1, \ldots, n\}$ participate, there is a characteristic function v which assigns a total cost for each coalition $S \in P$ with P comprising the set of all possible subsets of N. In our case, this cost is calculated by solving the corresponding optimization problem of the CVRPTW for the coalition S of customers (players). A cost allocation vector $z = (z_1, \ldots, z_j, \ldots, z_n)$ is such that allocates a cost for each player $j \in N$ such that:

$$\sum_{j \in N} z_j = v(N).$$

The equation above represents the "efficiency" condition in which the total cost of the grand coalition is split among its members according to z. Moreover, the "rationality" condition states that the sum of the allocated costs to each player of a coalition should be no bigger than the total cost of the coalition; e.g. no subset S of players can be formed where the cost resulting from a coalition with all of them is smaller than the total cost allocated to them according to z, namely:

$$\sum_{j \in S} z_j \leq v(S) \quad \forall S \in P.$$

In the context of a CVRPTW, a rational cost allocation vector would be one that allocates the cost to each customer such that no combination of customers can yield a smaller cost when served than the sum of their allocated values. The "individually rational condition" is a particular case for coalitions S of only one player, in which the cost allocated to it should not be bigger than its stand-alone cost: $z_j \leq v(\{j\})$. If a customer were to be assigned a cost bigger than its stand-alone cost it would not make any sense for the carriers to bring them into the pool of shared customers. Most importantly, as defined by game theory, the core in a cooperative game is the set of allocation vectors that satisfy the condition of rationality; that is:

$$\text{Core}(v) = \left\{ z \in \mathbb{R}^n : \sum_{j \in N} z_j = v(N), \sum_{j \in S} z_j \leq v(S) \quad \forall S \in P \right\}.$$

When an allocation vector is included in the core, it is said to provide "stability", given that there is no reason for players in the grand coalition to form a smaller coalition as this yields a worse outcome. The core is widely utilized as a measure of stability in research (Guajardo & Rönnqvist, 2016).

3.2.2. Shapley value allocation method

An answer to the profit(cost) allocation issue in cooperative games was introduced by Lloyd S. Shapley (Shapley, 1951, 1953). It distributes the total gains of the contributor players, assuming a non-zero contribution by all members of the coalition. The Shapley value is then calculated by accounting for all the possible orders a contributor may arrive to the coalition and computing each of their marginal contributions and average it over the total. It is based on a set of axioms which define the main characteristics a solution to the cost allocation problem should fulfill (Defryn, Sörensen, & Cornelissens, 2016). These axioms are the following:

- Symmetry: Interchangeable agents(customers) should receive the same share of the cost.
- Null player property: An agent who neither adds nor reduces the cost of any coalition should be allocated no cost.
- Efficiency: The sum of all costs allocated among all the agents should total the total coalition cost.
- Additivity: The sum of the cost allocated to agents a and b should be the same as that allocated to a third partner representing them.
- Individual rationality: The cost allocated to an agent cannot be larger than its stand-alone cost.

The allocation through the Shapley value to each player i can be calculated using 4 where the final term is the computation of the marginal contribution of player i to a certain sub-coalition S, being $v(S)$ the characteristic function which describes the worth of coalition S, e.g.: the total expected payoff S can obtain by cooperation. These marginal costs are calculated for each possible sub-coalition and then averaged by the weight that each possible combination of such sub-coalition may be reached.

$$\phi_i(N, v) = \frac{1}{N!} \sum_{S \subseteq N \setminus \{i\}} |S|!(|N| - |S| - 1)![v(S \cup \{i\}) - v(S)] \quad (4)$$

The following simple example depicts the calculation of the Shapley value. Suppose there are three players $\{1, 2, 3\}$ participating in a collaborative game with each of the possible sub-coalition resulting in the payoffs summarized in Table 3. Further, as we add players to the coalition in each of the 6 ways the grand coalition could be formed, each player in turn is going to add a marginal contribution as displayed in Table 4. Therefore, the expected marginal contribution and Shapley value for each player is simply the average value of these contributions.

In our scenario, we use the Shapley value to compute the cost each customer contributes to the total cost of serving them all together. That is, each customer is a player and we calculate the costs resulting from the objective function of the VRP for all the possible customer combinations.

Although the Shapley value defines a fair way of dividing the grand coalition's payment among its members, its result does not guarantee stability. This means some players might be better off when forming a different coalition than the grand coalition. However, the existence of a stable cost

Table 3: Payoffs of each sub-coalition for an example of a collaborative game.

S	null	{1}	{2}	{3}	{1,2}	{1,3}	{2,3}	{1,2,3}
v(S)	0	3	1	1	6	5	5	10

Table 4: Computation of the Shapley value for each player of a collaborative game.

Order	Marginal Contribution		
	1	2	3
(1,2,3)	3	3	4
(1,3,2)	3	5	2
(2,1,3)	5	1	4
(2,3,1)	5	1	4
(3,1,2)	4	5	1
(3,2,1)	5	4	1
Average	4.17	3.17	2.67

allocation is not assured and in a practical scenario, a company would only be able to accept whether to join the grand coalition or not, without any influence in the cost allocation (Vanovermeire et al., 2014). This is because the using the Shapley value method results in a solution which is "unique". Consequently, players do not have the information to challenge the allocated costs (Krajewska et al., 2008). Although, the non-emptiness of the core could be implemented as a way to verify the stability of the allocation solution and its effectiveness (Shi et al., 2020), we will not cover that aspect in our methodology. Moreover, the Shapley value derives in a cost allocation that is individually rational for a superadditive game such as the one concerning our work (Moulin, 1991).

As we have shown, the Shapley value of a participant is the expected marginal contribution of such player to a coalition picked randomly. From a practical perspective, the Shapley value has been applied in many real-world cases (Deidda et al., 2009; Littlechild & Owen, 1973; Moretti & Patrone, 2008; Moretti et al., 2007), being suggested as best practice in horizontal collaboration in logistics (Biermasz, 2012). Any cost allocation method that would steer away from the assumptions of the Shapley value method would also be valid for the application of this research, and would not modify the resulsts, to the extent that they remain a fair mechanism to divide the costs. Therefore, we leave the selection of the best allocation method outside of our work and use the Shapley value as the cost allocation mechanism for the remaining.

3.2.3. Approximation methods to the Shapley value: Random sampling

The drawback of the Shapley value being used as a cost allocation mechanism is the need of calculating the value of the characteristic function for every possible sub-coalition. This results in a very challenging computational effort that can rise exponentially to being impractical for even a low number of players. The number of possible subsets of a set consisting of n players is 2^n. Already for 30 players the amount of sub-coalitions rise over 1 billion. In particular, it is an NP-complete computing problem, therefore being too expensive computationally (Deng & Papadimitriou, 1994; Faigle & Kern, 1992).

An approximation method of the Shapley value using random sampling was first introduced by Castro et al. (2009). This procedure uses a random sample of the marginal vectors. Through sampling, statistical inference can be derived where it is impossible or impractical to obtain information for the entire population (Castro et al., 2009; Cochran, 1977). As shown by Castro et al. (2009) the estimations are efficient if the value of the characteristic function of any coalition can be calculated in polynomial time. The method consists of selecting a subset of orderings from all $n!$ possible orderings, computing the marginal contributions for all players and all orderings and approximating the Shapley value for each player by averaging the marginal contributions obtained. For the simple example introduced previously, a random sample of three orderings out of the total 6 gives the results shown in Table 5.

As expected this is very sensitive on the samples drawn for a low value of samples but becomes more accurate when increasing to a higher value of sampling (Castro et al., 2009). From the computation executed on problems of 10 customers, we conclude that with a sample size of 10 times the amount of customers an error no bigger than 3% is obtained for the computation of the Shapley value of a customer. Although this provides enough accuracy for our analysis, a better structured method is used as explained in the following section.

3.2.4. Structured Random Sampling

As shown in the example introduced and summarized in Tables 4 and 5, the calculation of the Shapley value through random sampling, yields an error which is partially due to the lack of ordering of players in the samples. Player 1 for example appears first in two of the samples and never appears last. These orderings prove to be very influential for the computing of the Shapley values.

A structured random sampling method was introduced by van Campen et al. (2018) which optimizes the random sam-

Table 5: Computation of the Shapley value through random sampling.

Order	Marginal Contribution		
	1	2	3
$(1, 2, 3)$	3	3	4
$(1, 3, 2)$	3	5	2
$(2, 1, 3)$	5	1	4
Average	3.67	3.00	3.33

pling method and reduces the resulting error. Their method seeks to ensure that each player is equally assigned to each position in the ordering. Therefore, each player's marginal contribution to a coalition of a defined size is calculated the same number of times and is equally distributed. To achieve this, a swapping mechanism is added to assign players in the right position in the orderings which are randomly selected. The total amount of orderings selected is divided into n subsets of length t in which each player is ordered r times in each position. The marginal contributions of the ordered player are then used to compute the Shapley value.

Considering the 3-person ($n = 3$) game example and the same $r = 3$ orderings randomly sampled before, we have therefore the size of each subset being $t = 1$. For each of the players now, each position is assigned t times in each of the subsets and their marginal contribution is calculated. This is summarized in Table 6. It can be observed that the values resulting from the use of structured random sampling are closer to the exact Shapley values than when using simple random sampling.

In their work van Campen et al. (2018) observe that, although the same number of marginal contributions per player is calculated, their method yields a small extra computational cost deriving from the swapping operation. Further, whereas random sampling is efficient, the structured method is shown by the authors not to be. Nonetheless, the structured random sampling method outperforms its simpler counterpart in terms of accuracy when it comes to approximating the Shapley values. On average the error in the Shapley value approximation is reduced by almost 30% with only a slight increase in average computation times. Consequently, we use structured random sampling in our work.

3.3. Regression analysis

For the statistical modelling part of our work, regression analysis will be conducted. Regression analysis is a set of statistical methods for determining the relationship between the dependent variable and the independent variable(s). These last are also called "explanatory" variables as the intention of the analysis is to find the degree to which each of these variables explains the behavior of the dependent one (Sykes, 1993).

Correlation is used which describes the association between two variables to the extent that a change in one is found next to a predictable change in the other. The correlation of two variables is calculated as the ratio between the covariance and the product of the standard deviations of each variable (Brown, 2016), that is:

$$\rho_{X,Y} = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y}$$

The correlation can be measured through the Pearson correlation coefficient which measures the linear correlation between two sets of data. Both the strength and direction of the association between the variables are expressed through the coefficient (Yule & Kendall, 1968). A positive coefficient indicates a direct relation whereas a negative, an inverse one. The strength is measured by how close the absolute value of the coefficient is close to 1, with 0 indicating no correlation (Dowdy et al., 2011).

With multiple linear regression, several independent variables x_i are assigned to one dependent variable y, which is expressed as a linear combination of the independent ones (Rawlings et al., 1998), that is:

$$y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n + \epsilon$$

where parameter α represents the intercept value of y when all independent variables are 0 and β_i is the coefficient to which each variable x_i is affected to explain the most of y. The regression error ϵ is added to account for the difference between the predicted and the actual observed value of y.

When performing linear regression, the ordinary least squares (OLS) method is one of the most used. It consists of a method for defining the unknown coefficients of the independent variables in a linear regression model by minimizing the sum of the squares of the regression errors.

When dealing with multiple linear regression it is of special importance to ensure that the underlying assumptions are satisfied. The most important are the following according to Osborne and Waters (2002):

- Variables are distributed normally
- Model is linear in parameters
- No multicollinearity, e.g. low correlation between independent variables
- Homoskedasticity; e.g. the variance of ϵ remains constant for all cases of x
- No serial-correlation between error terms

Table 6: Computation of the Shapley value through structured random sampling.

Subset	Order	Swap 1	Marginal Contribution 1	Swap 2	Marginal Contribution 2	Swap 3	Marginal Contribution 3
1	$(1,2,3)$	$(1,2,3)$	3	$(2,1,3)$	1	$(3,2,1)$	1
2	$(1,3,2)$	$(3,1,2)$	4	$(1,2,3)$	3	$(1,3,2)$	2
3	$(2,1,3)$	$(2,3,1)$	5	$(3,1,2)$	5	$(2,1,3)$	4
Average			4		3		2.83

To satisfy these assumptions, generally the data and features used in the model need to be reduced. Feature selection, in which a subset of relevant predictors is selected for use in a statistical model, can be used for this purpose. All these methods are applied in the statistical analysis of our work.

4. Data generation: Instances solving and cost allocation

In this section, we dedicate ourselves to the generation of data through the application of our methodology to different problems or instances. We start by solving simple CVRP without time window constraints so as to gain insights into the basic characteristics of the problem. Further, we continue with more well-known instances with the addition of time limitations. Lastly, we define the features that we use for the approximation of the costs allocated.

4.1. Analysis of small CVRP

For the analysis of simpler collaborative single-depot capacitated vehicle problems, we focus on the modelling of instances with 3 carriers who are assigned 10 customers randomly. 10 instances are modelled with customers distributed in a squared grid of dimension 10 as the one displayed in Figure 2a. For each of the instances, all combinations of customers are modelled and solved to optimization so as to obtain the exact Shapley value assigned to each customer. This value is illustrated in Figure 2b. It is worth noting that, for these simple instances and, with a similar share of customers for each carrier, the number of customers served by each vehicle is always between 3 and 4. Therefore, there is no value at this point in the research to use this measure in our analysis. We discuss this further in later sections.

Likewise, the independent CVRP for each of the carriers is solved and their individual cost is calculated. The savings each carrier gains from the collaboration are displayed in Table 7 where the total saving is the difference between the cost of serving each of their customers and the cost when summing the Shapley value assigned for the customers brought into the coalition. These savings are all positive regardless of whether all customers in each of the carriers' operations yield positive gains. Although the grand coalition might not be the best possible coalition for each carrier, it still yields a benefit when collaborating all together. Further, as explained in Section 3.2.2, due to the individual rationality of the Shapley value, the cost of a player always results in a lower value than its stand-alone cost.

For these instances, a linear regression model is created between the Shapley value assigned to each customer and the distance to the depot as the only dependent variable. The results, summarized in Table 8, show a high degree of correlation. We avoid imposing an intercept to our function given the nature of the problem so as to start from a base cost of 0 and only depend on the distance. The regression is showcased in Figure 3 where a clear fit is observed. Moreover, heteroskedasticity can also be distinguished from the increase in the variance of the errors as the distance increase. This suggests that the measure of distance becomes less of a predictor as customers are more distant from the depot where, given a bigger spread between customers, the limited capacity of the vehicles could play a bigger role. We deal with the problem of heteroskedasticity with robust regressions in the remaining of the work.

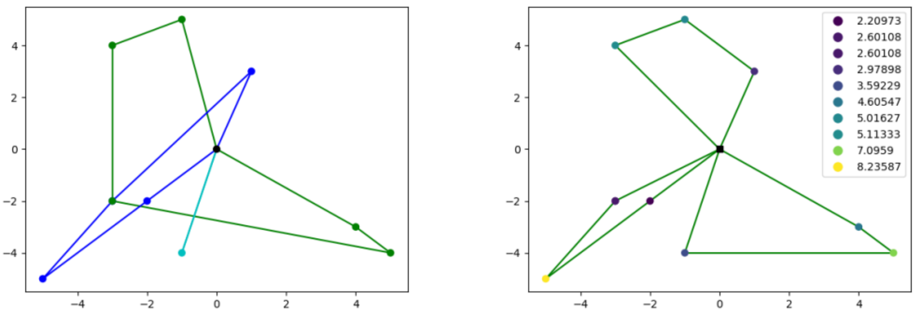
Finally, the result of the cost allocation is summarized in Figure 4 where the average cost for customers located in different placements is represented. It is noticeable how a simple rule based on distance can be applied to the solving of a much more complex logistics problem.

4.2. Solomon instances

In order to be able to benchmark the results obtained for the CVRPTW and ensure a correct implementation of the structured random sampling algorithm used for the allocation of costs, the use of well-known data instances is implemented, the Solomon instances (Solomon, 1987). Moreover, this allows for the further study of our results in consequent works.

Solomon instances are divided into different sets according to their characteristics. R and C sets in the data differ in that the R sets (R1 and R2) have the geographical data randomly generated, whereas the C sets (C1 and C2) are set in geographical clusters. For the RC1 and RC2 sets, a mix of random and clustered structures is used. The numbering relates to the limitation on the scheduling horizon, having smaller time windows and allowing a few customers per route in sets 1 and a long scheduling horizon with more customers in sets 2. The customer locations remain the same within all problems of one type (i.e., R, C and RC), mainly differing in the tightness of their time windows, that is the time window width. There exists also variance within the time window density, that is the percentage of customers with tight time windows, ranging from 25% to 100% (Solomon, 1987).

Solomon introduced problems of 100 customers where travel times match the euclidean distances between the ge-



(a) CVRP without collaboration of 3 partners (green, cyan and blue).

(b) CVRP with collaboration of 3 partners. Shapley value for each customer in color.

Figure 2: Instance of a collaborative vehicle routing problem with 3 partners and 10 customers.

Table 7: Total savings obtained for each carrier (A, B, C) when collaborating in a CVRP.

Instance	A	B	C
0	4.77	4.89	2.84
1	1.22	5.48	8.53
2	1.82	4.88	0.88
3	3.65	7.29	1.08
4	3.47	8.21	7.71
5	2.47	6.73	8.43
6	4.07	5.77	8.56
7	7.49	4.38	5.60
8	9.42	7.94	5.19
9	2.78	5.50	3.00
10	1.72	7.73	4.42

Table 8: Linear regression model summary with the cost per customer as the dependent variable and distance as the sole independent variable.

	coef	std err	t	$P > \|t\|$	[0.025	0.975]
distance	1.0162	0.02	51.11	0	0.977	1.056

ographical points. For computational reasons, we base our work on the instances of the 50 first customers in each of them where 100% of customers are assigned a time window constraint. The instances used in our analysis and their characteristics are summarized in Table 9.

Most importantly, the design of the instances highlights aims at addressing several variables that generally influence the routing and scheduling when solving CVRPTW problems such as geographical location, vehicle capacity, amount of time-constrained customers, and tightness and positioning of the time windows assigned. (Solomon, 1987).

4.3. Cost function: features and fit

With the aim of developing our cost function, different statistical models are applied to the data generated out of allocating the cost via our method using the Shapley value for each customer in Solomon's 50 customers instances. An example of such data can be observed in Figures 5 and 6 where the solution for instances R101.50 and R201.50 respectively are displayed. The vehicle routes are visible in conjunction with the computed Shapley values for each of the customers.

In the first example, having time windows which are more constraining, the vehicles route less customers in smaller loops. In contrast, in the latter, routes are more convoluted due to time constraints being more relaxed. This is also reflected in the spread of Shapley value which has a bigger variance and seems to depend more on distance in the first case compared to the second one. This example signals that short scheduling horizons, allowing for only a few customers to be serviced per route, yields a higher variance of the cost allocated for each customer.

An approximation to the cost function should try to resemble the real interactions of the CVRPTW as much as possible. For this reason, the definition of certain features should match the factors that affect this problem as discussed previously in Section 4.2. Therefore, different features have been

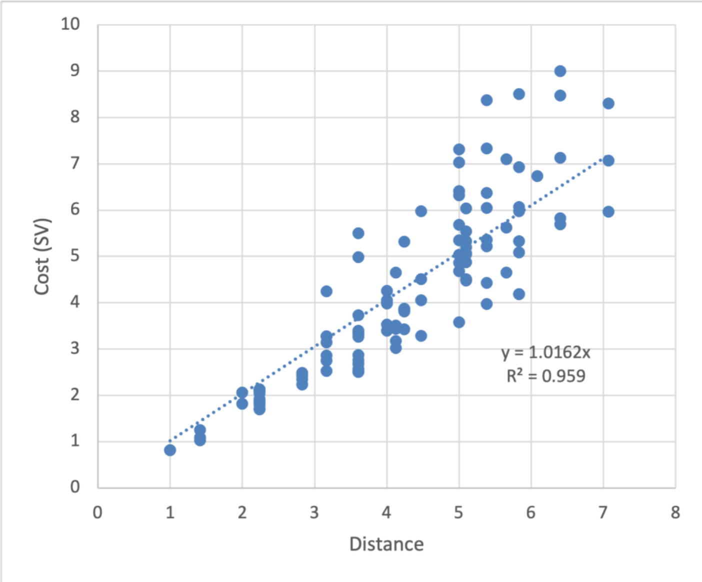


Figure 3: Relation between the cost allocated and the distance to the depot for simple CVRPs.

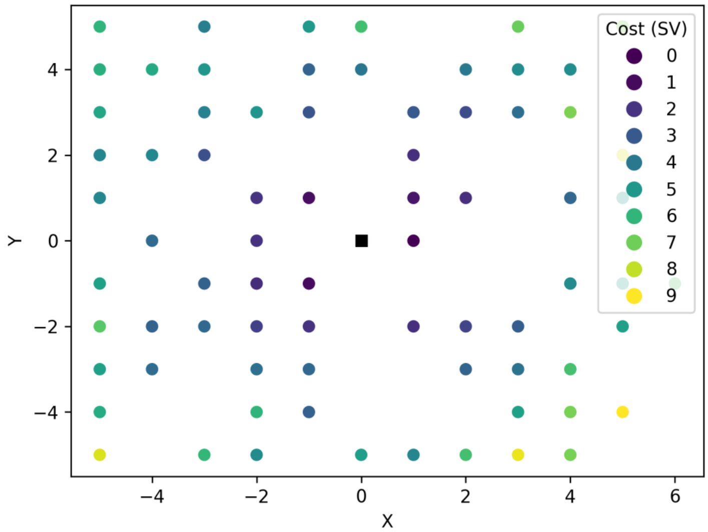


Figure 4: Cost allocated for customers in simple CVRPs.

defined following the three main dimensions that affect the cost allocation problem. In order to generalize their application and derive clearer conclusions from their coefficients, relative measures are used when possible.

Capacity-based features

The main parameter affecting the usage of vehicles is the demand of customers and the capacity of the vehicles. In order to represent the first effect in our function we introduce a feature relating the demand of each customer to the average demand of the population, that is:

$$\mathrm{CF}_1^i = \frac{d_i}{\frac{\sum_{i \in N} d_i}{|N|}}$$

Further the capacity of vehicles is reflected in the demand related to the vehicle capacity, e.g.:

$$\mathrm{CF}_2^i = \frac{d_i}{C}$$

where C is the vehicles' capacity.

Distance-based features

As observed in our modeled examples before, geographical distance appears as a main variable to take into consideration. Firstly, we introduce a feature comparing the distance of a customer i to the depot $i = 0$:

$$\mathrm{DF}_1^i = \frac{c_{i0}}{\frac{\sum_{i \in N} c_{i0}}{|N|}}$$

Table 9: Solomon instances used for the analysis and their characteristics.

Instance	Amount of Vehicles	Vehicles' Capacity	Average Demand	Average distance to depot	Average length of Time Windows	Service Time	Amount of customers w/ Time Window
C101.50	25	200	17.20	24.11	60.14	90	50
C201.50	25	700	17.20	26.79	160.00	90	50
R101.50	25	200	14.42	26.25	10.00	10	50
R201.50	25	1000	14.42	26.25	116.46	10	50
RC101.50	25	200	19.40	40.60	30.00	10	50
RC201.50	25	1000	19.40	40.60	120.00	10	50
C108.50	25	200	17.20	24.11	240.78	90	50
C208.50	25	700	17.20	26.79	640.00	90	50
R109.50	25	200	14.42	26.25	58.94	10	50
R209.50	25	1000	14.42	26.25	351.08	10	50
RC108.50	25	200	19.40	40.60	111.62	10	50
RC208.50	25	1000	19.40	40.60	472.90	10	50

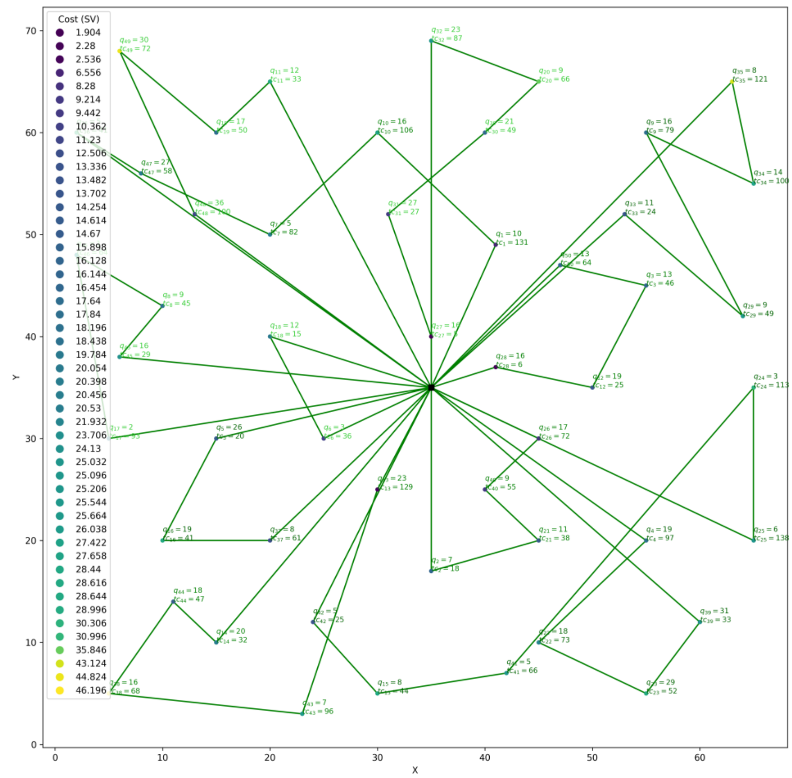


Figure 5: Solution to Solomon's instance R101.50 showing the computed cost (Shapley values) for each customer.

To note is that as all costs are taken using euclidean distances, the cost matrix is symmetrical, e.g.: $c_{ij} = c_{ji}$. Secondly, a measure of the degree of clustering is needed. Therefore the proximity to other customers is accounted for by measuring the number of points that fall within a certain distance. For our measure we use the average distance to the depot as the threshold for counting the near customers as shown in Figure 7. That is:

$$\mathrm{DF}_2^i = \|\{j \in J\}\| : d_{i,j} \in \left[0, \frac{\sum_{i \in N} c_{i0}}{|N|}\right)$$

Time-based features

In the case of the time dimension, the measure of the impact of the time windows' width is needed. With this objective, we introduce a feature in which we calculate the length of the time window as the difference between the earlier and latest service times and divide that by the average distance of customers to the depot.

$$\mathrm{TF}_1^i = \frac{b_i - a_i}{\frac{\sum_{i \in N} c_{i0}}{|N|}}$$

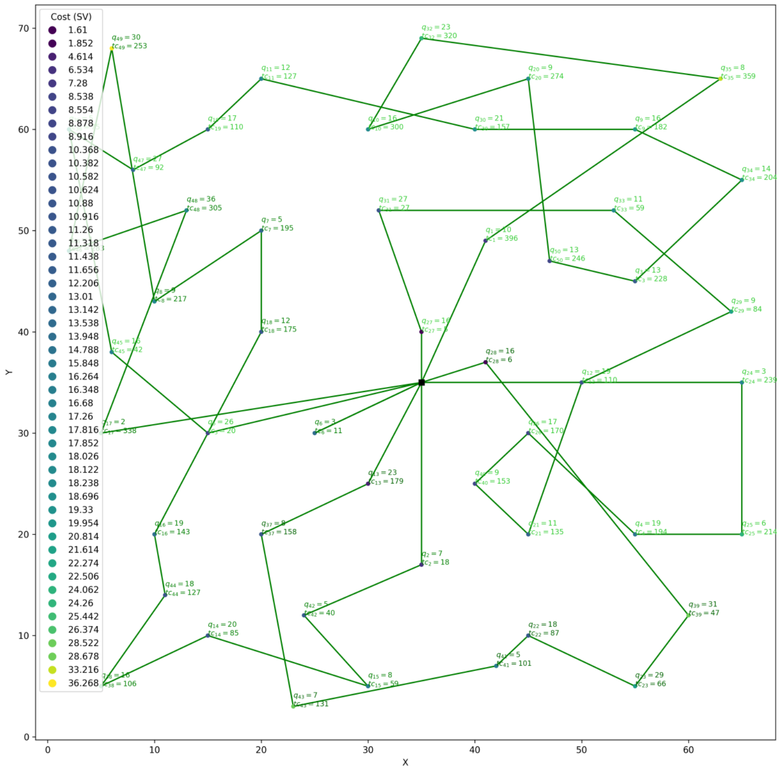


Figure 6: Solution to Solomon's instance R201.50 showing the computed cost (Shapley value) for each customer.

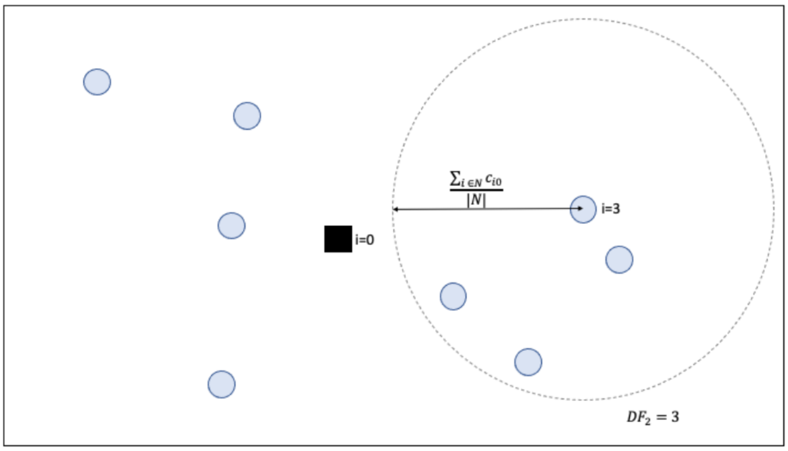


Figure 7: Example of computing the proximity customers for a customer i.

Further, as in the case of the distance between customers, the spread in the time windows is looked to be reflected in the time window overlap as defined in the time feature:

$$\mathrm{TF}_2^i = \|\{j \in J\}\| : (a_i, b_i) \cap \left(a_j, b_j\right) \neq \emptyset$$

Instance-specific features
Relying on the characteristics of Solomon instances which aim at characterizing the different dimensions affecting CVRPTW, we use certain instance-specific features which aid in the analysis of the costs and their distributions between instances themselves. For this purpose we separate instances under the following classification:

- Customers proximity: Clustered/Non-clustered customers
- Time: Tight/Non-tight Time Windows

Additionally, different modelled instances are compared using other characteristics which are not binary to the extent that they are not directly related to the Solomon instances' design criteria such as:

- Average customer demand

- Ratio of total demand served by vehicle capacity
- Average length of time windows
- Average number of customers per tour

5. Analysis and discussion

With the data generated for the instances used and the features calculated we follow with the statistical analysis in the following sections.

5.1. Results obtained and observations

The results obtained from applying our methodology are compiled in Appendix 1 where the characteristics of each node from Solomon instances in conjunction with the cost assigned in the grand coalition are shown.

As in Section 4.1, we begin by addressing the total savings of each carrier when collaborating which are summarized in Table 10 where the average value has been added for further study. We can again observe that all gains are positive, in line with the research findings explained in Section 2.1, and further enlarging the case for collaboration between carriers. Worth noting is that this is the total gains for each carrier (each of which has a similar share of assigned customers), which does not exclude the possibility of a customer bringing a negative gain when shared into the common pool of the joint operation. This is later explored in Section 5.4.

We can therefore observe the following:

Observation 1. The total savings calculated for each individual carrier when collaborating in a grand coalition are found to be positive in all instances.

Within this level of granularity, we want to evaluate the differences in the costs allocated and the corresponding savings obtained for the customers in each instance. Figure 8 depicts the average values obtained in instances split by clustering and the time window horizon. In Figure 8a we observe that the clustering of customers yields a smaller average cost per customer. Further, Figure 8b shows how having short time windows affect considerably the average cost of serving the customers. Interestingly, in both cases the savings do not show the same level of change.

To further study the instance-related features, Figure 9 compiles the same measures for average cost and savings compared to the different characteristics of the Solomon instances. Firstly, an increase in the average customer demand leads to an increase in the savings as shown in Figure 9a. Moreover, when analysing Figure 9b we observe no general trend for lower demands per truck capacity but notice an increase in the savings and costs when increasing past a certain threshold. We can relate the results from Figure 8b with Figure 9c where we observe that a decrease in the average time window length results in a smaller average cost allocated to each customer. Finally, with regard to the average number of customers per tour, Figure 9d shows that average costs decrease when this increases.

Throughout the results showed, a clear relation between costs and savings can be observed, pointing towards the fact that the same limitations that increase the cost of serving the pool of customers are the ones that can be avoided when collaborating.

Let us summarize the main observations in the following:

Observation 2. Customer clustering and relaxation of time windows yield a decrease in the average cost of serving each customer.

Observation 3. An increase in customer demand or a decrease in the vehicles' capacity leads to an increase in the costs associated with each customer.

Observation 4. As the total amount of tours to serve a certain amount of customers decreases, the average cost associated with each customer also does.

Observation 5. There is a strong correlation between the costs assigned to each customer when serving them in a stand-alone operation and the potential savings when acting in a collaborative environment.

Although always dealing with a collaborative scenario, the cost allocation deriving from the joint operation can be analyzed from a non-collaborative perspective. The grand coalition acts as a shared resources all-information central planning body. For the approximation of the cost allocation with the features defined, we start with the results obtained from the centrally planned cost allocation perspective. Later, we analyze the data resulting from expanding on the collaborative nature of the problem.

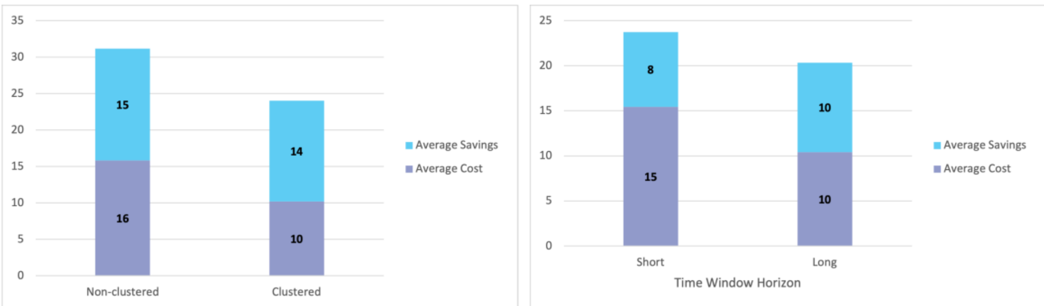
5.2. Model fit and feature selection

Our main aim is to conduct a model selection procedure to choose the features and interactions of them which yield the best fit to the cost allocated to customers through linear regression. The measure used to compare the performance of different models is under debate, which questions the "best" fit (Hansen & Yu, 2001). Nevertheless, we will focus on the Akaike (AIC) and Bayesian Information Criterion (BIC) and the adjusted R^2 value as means of evaluating our models. Furthermore, for simplicity of our solution and following the principle of parsimony, or Occam's razor, we will focus on finding the model with the least amount of features and interactions given similar explanatory power.

To start, a multiple ordinary least squares linear regression analysis is conducted using all features in the model as the independent variables and the calculated allocated cost (Shapley value) as the dependent one. Said analysis is performed with the data for all the combined instances shown in Table 9. At first, we conduct our analysis without including an intercept in the model, with the results being presented in Figure 10. By applying a Breusch-Pagan test to the data, the null hypothesis of homoskedasticity is discarded therefore the regression results show the scores for t-tests using heteroskedasticity-robust standard errors. The model shows a high correlation with a value of R^2 of 0.855. Regarding the distance metrics, from the values of the regression coefficients, we can infer both a positive statistically significant

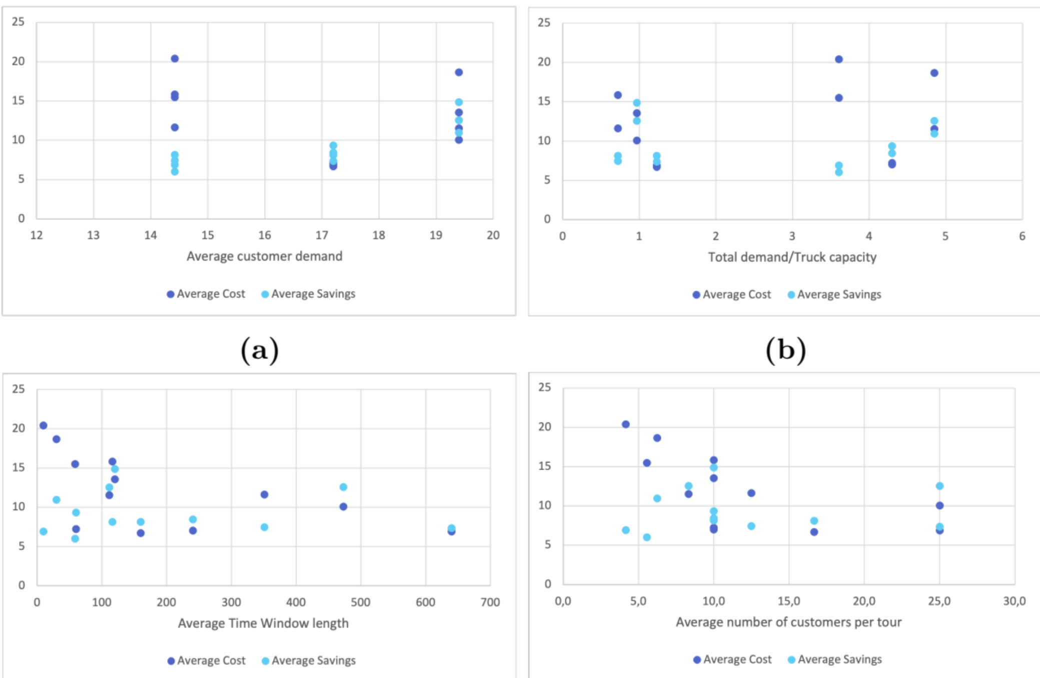
Table 10: Total savings for each carrier (A, B, C) in the modelled CVRPTW Solomon instances.

Instance	A	B	C	Average
C101.50	150.75	161.57	154.59	155.64
C201.50	135.59	113.51	156.80	135.30
R101.50	120.20	123.43	101.63	115.09
R201.50	145.82	131.90	128.99	135.57
RC101.50	228.59	152.04	166.80	182.48
RC201.50	263.92	251.12	228.17	247.74
C108.50	133.97	151.41	136.62	140.66
C208.50	127.60	102.38	137.47	122.49
R109.50	78.19	92.84	129.13	100.05
R209.50	112.12	124.45	136.17	124.25
RC108.50	219.75	204.91	202.62	209.09
RC208.50	210.48	213.47	204.14	209.36



(a) (b)

Figure 8: Average costs and gains for the modelled Solomon instances split by clustering and TW horizon.



(a) (b)

(c) (d)

Figure 9: Average costs and savings for the modelled Solomon instances in relation to different instance characteristics.

correlation between the cost and the relative distance to the depot (DF1) and a negative one with the amount of proximity points (DF2). This confirms the results previously discussed in Section 5.1 in Observation 2. With regard to the time features, the negative correlation of the relative time window length (TF1) signals a decrease in the cost when having less constraining time windows which also is in line with the previous observations. Further, a positive coefficient for the time window overlap (TF2) would be against the common sense of pricing time windows and will be explained through further analysis of different models. None of the capacity constraints appear to have any statistical significance given by the p-values of capacity-related features which we analyze later.

To further explore the correlation between the variables let us analyze the relationship between each of them. Figure 11 shows a scatter plot for each of the variable pairs in the regression model. A grid of Axes is displayed such that each variable in the data will be shared across the y-axes across a single row and the x-axes across a single column. The diagonal plots show a distribution plot to represent the marginal distribution of each variable. Through an inspection of the scatter plots for the features in each dimension a clear correlation can be observed, in particular within distance-dependent and time-dependent features. This is further confirmed when calculating the correlation values as shown in Figure 12. As observed, the correlation between DF1 and DF2 is negative which means that having more points in the vicinity relates to a decrease in the distance to the depot or vice versa. Further, the correlation between features TF1 and TF2 is positive corresponding to a increase in the time window overlap when increasing the gap between the beginning and end of the allowed time for service at the customer.

As observed in the relation between the features in each dimension, and due to over-redundancy, the results obtained signal the need of relying on one main factor for each of the dimensions specified (Capacity, Distance, Time). A model selection is then conducted to choose the best fit with the measures specified before. All the combinations of features are tested and their AIC, BIC and adjusted R^2 values are calculated. It is found that features DF1, TF1 and CF2 are the ones that yield a better explanation of the dependent variable remaining one of the relatively lowest models regarding the AIC and BIC scores. It is reasonable to believe that the inclusion of CF2 is related to the addition of the maximal vehicle capacity as a variable to the model. Further, of the proximity measures included for distance and time windows none remain as the selected by the model. As we see later, this is very much related to the degree customers are clustered and time-constrained. The results for the regression with only these features are showcased in Table 11. The feature CF2 appears to have less statistical significance but has a negative coefficient which is in line with Observation 2. It is worth noting that the addition of an intercept in our model yields a lower explanatory effect over the independent variable given its R^2 value but has a higher statistical significance for all the variables including the intercept and achieves a lower AIC and BIC value as displayed in the regression results in Figure 13. The addition of an intercept would signify a base cost allocated to customers in any case which would be a simple base pricing strategy from which to start.

From the two linear regression models, with or without intercept, we can derive two main strategies to price customers. Firstly, one where no base price is defined and through the implementation of features pricing the dimensions which define each customer (demand, time window, distance) the final cost is greatly described. Secondly, assuming a basic price for all customers (intercept) and pricing the features differently resulting in reduced descriptive power of each feature on the final cost allocated to each customer. We will further study the pricing of the different features in the following section.

As a whole, from all our regression models we can observe the following:

Observation 6. The cost allocated to customers in vehicle routing problems with time windows can be estimated with a few defined features related to customers' location, time window, demand and vehicle capacity.

5.3. Customer pricing

The practical application of our work is focused on being able to price each of the characteristics that comprise the logistics problem of serving different customers whose geographical location, time demands and capacity vary. As defined in our cost-approximation function, the features introduced allow us to represent each of these dimensions in our model. None-withstanding, we want to study different ways of approaching these characteristics. As addressed in Section 4.2 the variety in the Solomon instances allow for the assessment of the results of our cost approximation when dealing both with clustered customers vs non-clustered ones and the time horizon restrictions. In this regard, we will study the differences in our regression models when dealing with each condition.

5.3.1. Clustered vs non-clustered customers

For this comparison, instances with non-clustered customers which are located geographically in a random fashion (R101.50, R201.50, R109.50, R209.50) are compared with their clustered counterparts in which customers are in groups of close proximity to each other (C101.50, C201.50, C108.50, C208.50). Instances RC, being a mix of both, are left outside of the analysis. The results of the linear regression for both cases are compiled in Tables 12 and 13 where CF1 has been dropped for being highly correlated to CF2. The explanatory power of the models gives an adjusted R^2 value of 0.624 for the clustered customers and 0.702 for the randomly located ones. It can be observed that for the non-clustered customers the intercept for the cost is not statistically relevant with only the relative distance to the warehouse being the relevant distance feature. In contrast, when dealing with instances of clustered customers, the intercept

```
                            OLS Regression Results
=======================================================================================
Dep. Variable:                   Cost   R-squared (uncentered):                   0.855
Model:                            OLS   Adj. R-squared (uncentered):              0.853
Method:                 Least Squares   F-statistic:                              686.2
                                        Prob (F-statistic):                   2.63e-263
                                        Log-Likelihood:                         -1859.1
No. Observations:                 600   AIC:                                      3730.
Df Residuals:                     594   BIC:                                      3757.
Df Model:                           6
Covariance Type:                  HC1
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
DF1           14.7125      0.673     21.848      0.000      13.393      16.032
DF2           -0.0542      0.023     -2.394      0.017      -0.099      -0.010
CF1            0.4697      0.503      0.933      0.351      -0.517       1.456
CF2          -10.7873      6.737     -1.601      0.109     -23.993       2.418
TF1           -0.3918      0.035    -11.219      0.000      -0.460      -0.323
TF2            0.0334      0.014      2.424      0.015       0.006       0.060
==============================================================================
Omnibus:                       45.913   Durbin-Watson:                   0.952
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               64.946
Skew:                           0.589   Prob(JB):                     7.89e-15
Kurtosis:                       4.100   Cond. No.                         936.
==============================================================================
```

Figure 10: Ordinate least squares regression results for all instances with all features in the model.

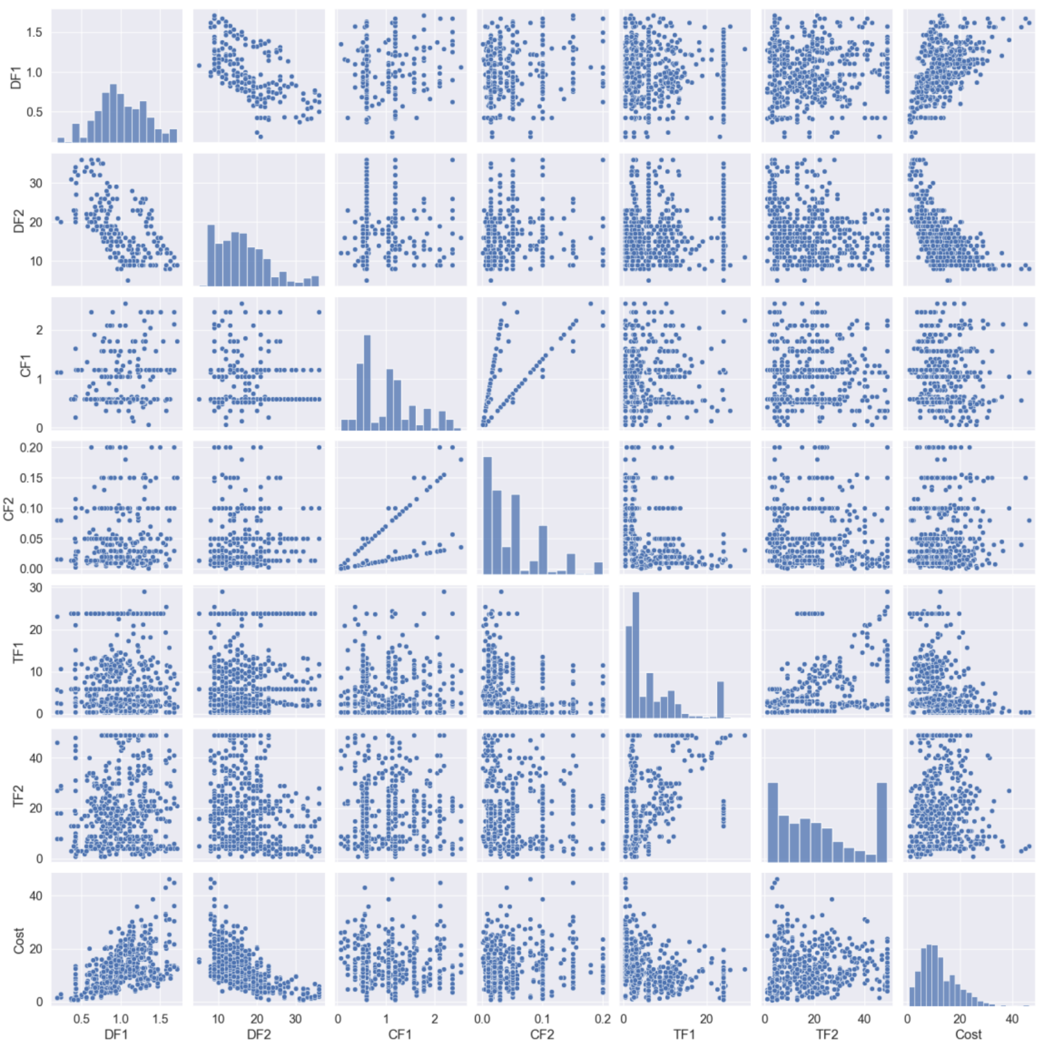


Figure 11: Scatter plots of all the pairs of variables in the model. The variables' distributions are shown in the diagonal plots.

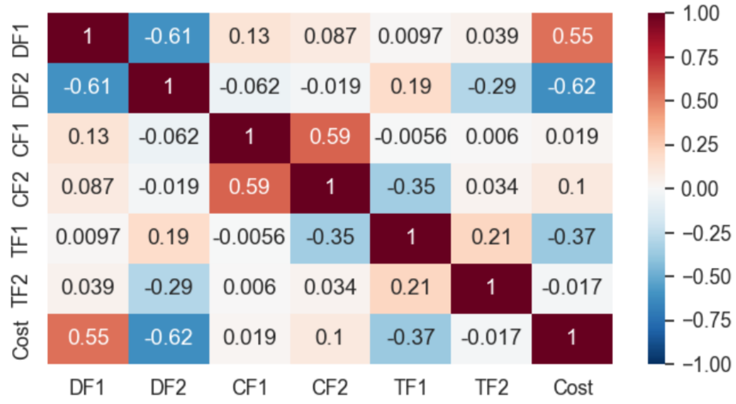


Figure 12: Correlation values between each feature in the model.

Table 11: Linear regression model summary with the cost (Shapley value) as the dependent variable and the selected features as independent variables.

	coef	std err	z	$P > \|z\|$	[0.025	0.975]
DF1	14.933	0.492	30.380	0.000	13.970	15.897
TF1	-0.388	0.033	-11.893	0.000	-0.452	-0.324
CF2	-8.322	4.829	-1.723	0.085	-17.787	1.143

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                   Cost   R-squared:                       0.449
Model:                            OLS   Adj. R-squared:                  0.447
Method:                 Least Squares   F-statistic:                     106.8
                                        Prob (F-statistic):           2.43e-55
                                        Log-Likelihood:                -1859.3
No. Observations:                 600   AIC:                             3727.
Df Residuals:                     596   BIC:                             3744.
Df Model:                           3
Covariance Type:                  HC1
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept      2.6905      0.790      3.405      0.001       1.142       4.239
DF1           12.9943      0.854     15.221      0.000      11.321      14.667
TF1           -0.4278      0.031    -14.011      0.000      -0.488      -0.368
CF2          -13.7909      5.104     -2.702      0.007     -23.794      -3.788
==============================================================================
Omnibus:                       58.953   Durbin-Watson:                   0.977
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               90.781
Skew:                           0.682   Prob(JB):                     1.94e-20
Kurtosis:                       4.331   Cond. No.                         219.
==============================================================================
```

Figure 13: Regression results for the selected model features with an added intercept.

becomes relevant in conjunction with the two distance-based features. In particular, the addition of DF2 which considers the proximity to other customers affects the cost negatively. Further, for the randomly located customers, the time features related to the length and overlap of time windows appear to be statistically significant and affect both the cost negatively. This would signify that time windows have a major impact on the cost of serving customers when these are not grouped in clusters of close proximity. Moreover, the difference in the coefficient for DF1 of more than 10 signifies a significantly bigger impact on the distances in the case of non-clustered customers compared to the clustered ones. All in all, our results clearly show how the grouping of customers in clusters can have a clear effect on the variables which mainly affect the allocation of costs.

Further, having the geographical location as the only feature we would like to assess the pricing of certain regions or locations for our customers. In this regard, we compile all the customers from the instances mentioned above so as to show how distance to the depot, which translates into time in our model, affects cost allocation. Figure 15 shows the case of customers with random locations where zoning related to the distance to the warehouse is applied. In the case of clustered customers we summarize the customers of in-

Table 12: Linear regression results for the non-clustered customers.

	coef	std err	z	$P > \|z\|$	[0.025	0.975]
const	10.011	5.494	1.822	0.068	-0.757	20.779
DF1	14.666	2.351	6.238	0.000	10.058	19.274
DF2	-0.282	0.206	-1.366	0.172	-0.687	0.123
CF2	-4.769	9.865	-0.483	0.629	-24.104	14.566
TF1	-0.328	0.076	-4.296	0.000	-0.477	-0.178
TF2	-0.121	0.030	-4.057	0.000	-0.180	-0.063

Table 13: Linear regression results for the clustered customers.

	coef	std err	z	$P > \|z\|$	[0.025	0.975]
const	7.230	1.360	5.318	0.000	4.565	9.895
DF1	4.272	0.537	7.952	0.000	3.219	5.325
DF2	-0.215	0.036	-6.019	0.000	-0.285	-0.145
CF2	1.374	3.472	0.396	0.692	-5.431	8.179
TF1	-0.033	0.030	-1.121	0.262	-0.091	0.025
TF2	0.029	0.024	1.201	0.230	-0.018	0.077

stances RC101.50, RC201.50, RC108.50 and RC 208.50 in their respective clusters and then calculate an average cost per cluster which is represented in Figure 14. Both figures exemplify how different zoning rules can be applied to better adapt to the different structures of customer locations.

5.3.2. Tightness of time windows

As in the case of distance, we proceed to analyze the effect of time in our problem by comparing instances which have a shorter time-horizon (e.g. R/C/RC 101.50) to those which a longer one (e.g. R/C/RC/201.50). For the case with tighter time restrictions, it can be observed that the significant features are only the relative distance to the depot (DF1) and both time features. The limitation that the time constraint produces, makes the overlap of time windows (TF2) a significant feature in contrast with the case with longer time horizons. Moreover, in the case with more relaxed time constraints the intercept becomes statistically significant yielding a basic cost as a base from which then the first distance measure and the relative length of the time windows both add and subtract respectively to the cost allocated for each customer. Further, the clustering of customers has a statistically significant correlation in this scenario as is the case with the capacity.

In general, we observe that short scheduling horizons, allowing for only a few customers to be serviced per route, yield a higher variance of the cost allocated to each customer.

To summarize, the individual regression models of each of the different set of instances has shown how the pricing strategy of customers depends greatly on the configuration of them. It can be derived that the classification of customer configurations among the categories used results in a better understanding of how to divide the costs when serving them.

5.4. Collaborative gains analysis

One of the main areas of focus of our work is the cooperation aspect of the logistics carriers in this problem. Given that we are dealing with a collaborative scenario, we are interested in evaluating the allocation of costs of each carrier when operating individually compared to that when in cooperation. So far, we have allocated costs based on a joint logistic operation in which we have assumed that all carriers share their entire pool of customers and vehicle resources so as to achieve the most efficient service from a logistics perspective. There is however a reluctance of carriers to share all their customers' information given that this is from where their business extracts great value.

As explained, for each of the instances modelled we have assigned each customer in the pool to one of three carriers. We therefore utilize our methodology within each carrier's operation solving the CVRPTW and using structured random sampling to calculate the Shapley value and allocate the cost of each customer in a partner stand-alone operation. These results can be found in Appendix 1 under the carrier's assignment and the stand-alone cost.

Moreover, for each of the customers of the competitors, a cost is calculated considering what would be the cost allocated to that customer if the entire operation was to be delivered to an identical pool of customers with the only addition of this extra one. This is represented in Figure 16. In this scenario, the mechanism of an auction is being reproduced where a carrier would offer a customer to the pool of shared customers. In 5.9a carrier 1 would offer a customer for which it has a high cost allocated. Secondly, in 5.9b the other carriers calculate their cost when adding this customer to their service. Finally, the carrier with the lowest servicing cost wins the auction given that he can pay a higher price because he can include the customer with the lower marginal cost. This results in the transfer of this customer from one operator to the other. Considering this, we calculate for each

Cluster	Distance to Depot	Cost
1	37.8	12.2
2	36.5	11.5
3	36.1	12.4
4	40.8	13.0
5	51.0	18.1

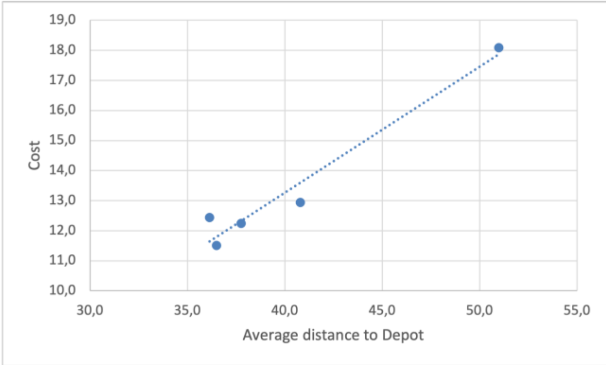


Figure 14: Cost assigned for each customer cluster in clustered instances.

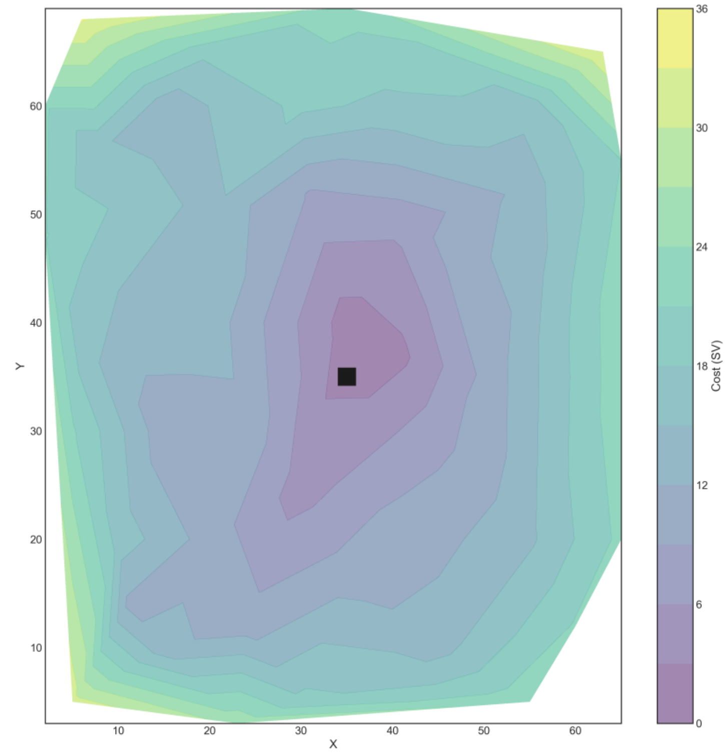


Figure 15: Distribution of cost for customers in instances with randomly chosen locations.

Table 14: Linear regression results for the instances with a shorter time horizon.

	coef	std err	z	$P > \|z\|$	[0.025	0.975]
const	5.490	3.071	1.788	0.074	-0.529	11.508
DF1	16.134	2.187	7.378	0.000	11.847	20.420
DF2	0.029	0.087	0.335	0.737	-0.141	0.200
CF2	-16.001	8.931	-1.792	0.073	-33.505	1.504
TF1	-5.495	0.566	-9.712	0.000	-6.604	-4.386
TF2	0.126	0.057	2.191	0.028	0.013	0.239

carrier the Shapley value for each customer of the competitors. The lowest cost from the computation of other carriers' operations is presented in the last column in the table found in Appendix 1. Our calculations are not completely representative of the entire auctioning process because they only consider the first iteration of it. This proves to be sufficient for our analysis.

We follow with the analysis of two scenarios. In the first

Table 15: Linear regression results for the instances with a longer time horizon.

	coef	std err	z	$P > \|z\|$	[0.025	0.975]
const	17.964	2.938	6.115	0.000	12.207	23.722
DF1	7.918	1.593	4.969	0.000	4.795	11.041
DF2	-0.402	0.067	-5.972	0.000	-0.533	-0.270
CF2	-98.750	35.031	-2.819	0.005	-167.410	-30.091
TF1	-1.096	0.418	-2.622	0.009	-1.915	-0.277
TF2	-0.042	0.090	-0.470	0.638	-0.219	0.134

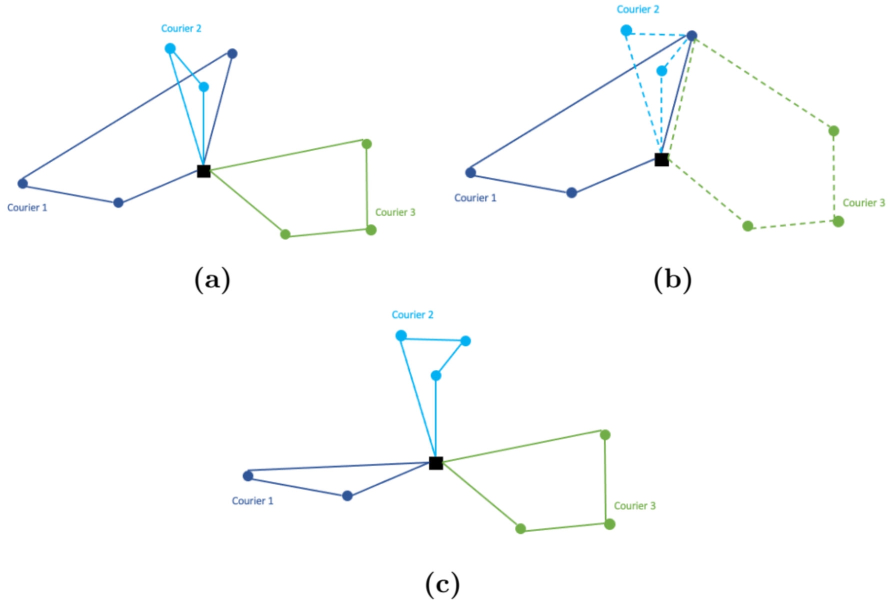


Figure 16: Representation of an auction mechanism for the sharing of customers in a collaborative vehicle routing problem.

one, we compare the stand-alone operation of each carrier with the grand coalition (central plan) where all customers are shared. For the second scenario, we consider the aforementioned auction mechanism in which we compare the cost of every customer in each carriers' operation to that of their competitors.

5.4.1. Central plan comparison

The aim is to derive insights into which customers become detrimental when being shared which defeats the purpose of sharing the information in the search of reducing costs. Accordingly, we calculate customers' Shapley value in carriers' stand-alone operation and in a grand coalition. We then classify customers who yield a negative/positive gaining so as to perform a logistic regression with the variables in the data. A dummy variable is used with a value of 1 for customers with a positive gain when shared and 0 otherwise. The aim is to have a prediction of which customers should be shared and which should be kept. Consistently with the rest of our work, the instances mentioned in Section 4.2 are used. Of the 600 customers, 33 yield a negative gain. From our statistical analysis, only the relative distance yields any statistical significance in the explanation of the dependent variable. The results for the regression when using only this feature are condensed in Figure 17.

Further, the gains for each customer when cooperating are also regressed with the features constructed so far. The results are showcased in Table 16 where the effect of both distance features are found to have a significant correlation.

Apart from the significant descriptive value of the distance features, there is no particular difference found for the 33 customers for which gains are negative. This constitutes a paradox, given that the cost for these customers is higher in the collaborative scenario than when in a partner stand-alone undertaking, but still carriers benefit from the collaboration as a whole when summing all customers. There is, therefore, a contradiction between the cost-allocating problem and the logistics problem which represents an area of potential future work.

In general, we found a very strong correlation between the costs allocated to customers when carriers act alone and those when doing so in a collaborative scenario. Figure 18 shows this relationship which yields a high value of R^2, confirming Observation 5.

5.4.2. Auction scenario

Analogously, the same procedure from the central plan scenario is applied to the auction one. Of the 600 customers

```
                           Logit Regression Results
==============================================================================
Dep. Variable:          sell_customer   No. Observations:                  600
Model:                          Logit   Df Residuals:                      599
Method:                           MLE   Df Model:                            0
                                        Pseudo R-squ.:                -0.01568
                                        Log-Likelihood:                -129.79
converged:                       True   LL-Null:                       -127.79
Covariance Type:            nonrobust   LLR p-value:                       nan
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
DF1            3.0327      0.207     14.621      0.000       2.626       3.439
==============================================================================
......
```

Figure 17: Logistic regression result when predicting a customer's gain in collaboration.

Table 16: Linear regression results for the gains obtained for each customers when carriers collaborate.

	coef	std err	z	$P > \|z\|$	[0.025	0.975]
const	8.258	1.988	4.153	0.000	4.361	12.155
DF1	6.570	1.153	5.700	0.000	4.311	8.830
DF2	-0.370	0.062	-5.931	0.000	-0.492	-0.247
CF2	11.034	9.540	1.157	0.247	-7.664	29.732
TF1	0.044	0.043	1.017	0.309	-0.040	0.127
TF2	-0.007	0.021	-0.328	0.743	-0.047	0.034

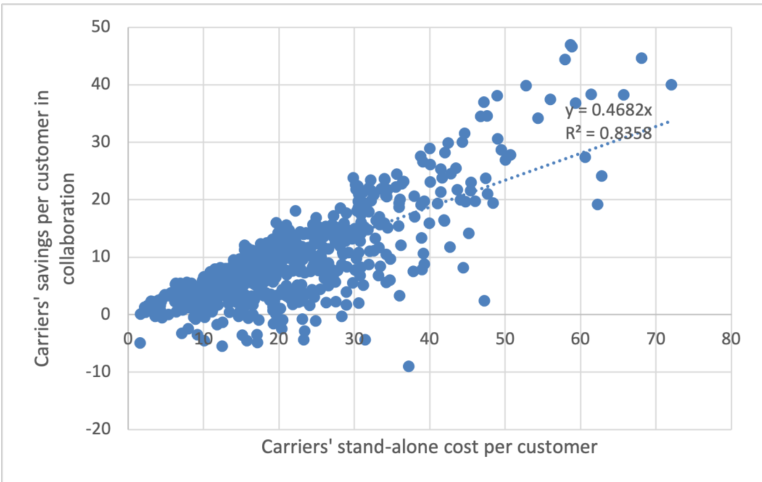


Figure 18: Correlation between savings in collaboration and costs in stand-alone allocated to customers.

studied, 167 yield a negative gain when comparing it to the lowest servicing cost from other carriers; that is, only 167 customers are serviced with the lowest cost by the carrier they have been assigned to. We follow with the same logistic regression as with the case before which only has a significant correlation with DF1 with a coefficient of 0.90. Moreover, a linear regression model is constructed with the defined features as independent variables and the gains by auctioning as the dependent one. The results from the OLS regression is shown in Table 17. Although having a low explanatory power with an adjusted R^2 value of 0.069 , the intercept has a significant statistical value. Noticeably, the only other feature with statistical significance is DF2 measuring the degree of clustering of customers. This represents that customers which are closer to other customers yield a lower cost for other carriers than the one serving them.

Given the outcome observed for this scenario, it is sensible to expect all total savings of carriers when collaborating to be positive, as our results in Section 5.1. This is explained in the fact that the great majority of customers can be serviced with lesser cost by a different carrier than the one to which they have been assigned. In practical scenarios, decentral collaboration can be exercised for the sharing of selective customers which are found to bring the biggest cost. In our results, we find that the average cost allocated in a standalone operation to customers that are cheaper to service by other carriers is 23.84 compared to 15.32 for those which are not. This represents a significant difference of 56% bigger cost.

Table 17: Linear regression results for cost of customers compared to other carriers' cost.

	coef	std err	z	$P > \|z\|$	[0.025	0.975]
const	7.561	2.288	3.305	0.001	3.077	12.045
DF1	1.669	1.371	1.218	0.223	-1.018	4.356
DF2	-0.292	0.069	-4.213	0.000	-0.428	-0.156
CF2	15.911	9.921	1.604	0.109	-3.533	35.355
TF1	0.011	0.047	0.236	0.813	-0.082	0.104
TF2	-0.027	0.023	-1.179	0.238	-0.071	0.018

6. Conclusion and recommendations

This study explores how cost allocation to customers served in Capacitated Vehicle Routing Problems with Time Windows can be approximated through the use of distinct features which relate to the specific constraints of the problem. From a methodology perspective, in this exploration a computationally efficient method to allocate costs and test different regression models so as to find the best fit is found. This is done by combining different methods both to solve efficiently CVRPTW with a metaheuristic solver and to allocate costs accurately with a structured random sampling method to approximate the Shapley value.

In the case of simple CVRP the measure of distance is found to achieve a high explanatory power over the cost allocated to customers when dealing with less capacitated instances and low distances. Additionally, the distribution, and therefore the uncertainty, of the allocated costs grow with increasing distances where demand and vehicles' capacity limit the problem increasingly.

As for the CVRPTW, the main characteristics of the well-known Solomon instances modelled, such as the clustering, capacity and time window length, allow for the observation of how these variables and limitations of the problem affect the cost of the logistics operation. These observations are confirmed by analysing the cost allocated to each customer individually. Customer clustering and relaxation of time windows is found to decrease costs. Moreover, the ratio between customer demand and vehicle capacity and the amount of vehicles needed in general lead to increased costs. On the whole, it is found that costs are strongly correlated to potential savings from collaboration.

In order to approximate the calculated costs through a linear function, this study compiles different features to condense the variables affecting the problem. Consequently, through model fitting and feature selection, the most important features are defined and a cost prediction model is created. Later, the differences in models when working with different instances are assessed. Most importantly, it is concluded that the specifics of the CVRPTW, and the extent to which the constraints limit them, are decisive when approximating the cost allocation methods through a formula. Nonetheless, from the analysis of the instances in this study it is inferred that a high degree of accuracy can be achieved on the regression of costs allocated with only a small amount of features.

All in all, the classification of practical problems according to their constraints can result in great value to determine which factors are of utmost relevance. In the case of clustered customers, the relative distance to the depot and the measure of clustering are of strong importance. In contrast, with non-clustered customers, time windows gain major importance due to customers not being in proximity to each other. Further, different kinds of zoning suit better each kind of problem. Radial zoning adjusts better for non-clustered customers, whereas local zoning does so for clustered ones. With respect to time restrictions, it is observed that short scheduling horizons result in a higher variance of the costs allocated with a dependency on time window length. In conclusion, cost-allocating mechanisms and resulting vectors benefit from considering the different types of categorizations analyzed throughout our work; e.g. when collaborating, carriers can align on the clustering of neighborhoods and therefore choose to participate in only selected ones.

Two main pricing strategies are found, with their scope depending on the problem constraints: With or without a base price. It is derived that the classification of customer configurations among the categories used results in a better understanding on which strategy to utilize.

Further, this work assesses the logistic problem in the context of collaboration. The gains obtained by carriers when acting in a joint operation are determined and the impact of constraints on their costs and potential savings analyzed. The total sum of gains obtained by carriers when acting jointly is found to be positive in all instances studied. In general, there is a great degree of correlation found between the cost of carriers when acting in a stand-alone and the gains from collaborating. Nonetheless, a small fraction of customers is found where the grand coalition assigns a bigger cost than the stand-alone cost obtained with one carrier. An area of further study is identified where the underlying reasons for this result could be explored and whether or not the use of the Shapley value, lacking stability from a game theory application, could be improved as a cost allocation method. Furthermore, from a decentralized collaboration perspective, customers which can be served by other partners more effectively are found to be 56% more costly than the rest in each carrier's stand-alone operation signaling for the need of further developing this area of study.

This work is focused on deterministic scenarios. Potential paths of future research could be explored with stochastic demands or uncertainties in other variables. In contrast with static assumptions in this research, dynamic changes with multiple periods could also be analyzed. Concerning the collaborative environment studied, auction scenarios represent a valuable area of further exploration. Moreover, the assumptions of having central planning and one main depot can be expanded into multi-depot and selective customer-sharing problems. To complement this study, the inclusion of more carriers and vertical integration could better enrich this research, with multi-modal logistics adding to the possibilities. In addition, the cost allocation mechanisms utilized could be applied in areas related to environmental impacts and their distribution. Finally, this work does not address the cost of sharing information and the extent to which it needs to be achieved for reasonable collaboration with regard to profits.

References

Agarwal, R., & Ergun, Ö. (2010). Network design and allocation mechanisms for carrier alliances in liner shipping. *Operations Research*, *58*(6), 1726–1742.

Akkerman, F., & Mes, M. (2022). Distance approximation to support customer selection in vehicle routing problems. *Annals of operations research*, 1–29.

Alfonso, V., Boar, C., Frost, J., Gambacorta, L., & Liu, J. (2021). E-commerce in the pandemic and beyond. *BIS Bulletin*, *36*(9), 1–9.

Anily, S., & Haviv, M. (2007). The cost allocation problem for the first order interaction joint replenishment model. *Operations Research*, *55*(2), 292–302.

Audy, J.-F., D'Amours, S., & Rönnqvist, M. (2012). An empirical study on coalition formation and cost/savings allocation. *International Journal of Production Economics*, *136*(1), 13–27.

Audy, J.-F., D'Amours, S., & Rousseau, L.-M. (2011). Cost allocation in the establishment of a collaborative transportation agreement—an application in the furniture industry. *Journal of the Operational Research Society*, *62*(6), 960–970.

Ballot, E., & Fontane, F. (2010). Reducing transportation CO2 emissions through pooling of supply networks: perspectives from a case study in French retail chains. *Production Planning & Control*, *21*(6), 640–650.

Berbeglia, G., Cordeau, J.-F., Gribkovskaia, I., & Laporte, G. (2007). Static pickup and delivery problems: a classification scheme and survey. *TOP*, *15*, 1–31.

Berger, S., & Bierwirth, C. (2010). Solutions to the request reassignment problem in collaborative carrier networks. *Transportation Research Part E: Logistics and Transportation Review*, *46*(5), 627–638.

Biermasz, J. (2012). *Report on the legal framework for horizontal collaboration in the supply chain* (tech. rep.).

Bratton, T. W., Mentzer, J. T., Foggin, J. H., Quinn, F. J., & Golicic, S. L. (2000). Supply chain collaboration: the enablers, impediments, and benefits. *Council of Logistics Management Fall MeetingCouncil of Logistics Management*.

Bräysy, O., & Gendreau, M. (2005). Vehicle routing problem with time windows, Part II: Metaheuristics. *Transportation science*, *39*(1), 119–139.

Bretzke, W.-R. (2013). Global urbanization: a major challenge for logistics. *Logistics Research*, *6*(2-3), 57–62.

Brown, R. (2016). Fundamentals of Correlation and Regression.

Caprara, A., & Letchford, A. N. (2010). New techniques for cost sharing in combinatorial optimization games. *Mathematical Programming*, *124*, 93–118.

Castro, J., Gómez, D., & Tejada, J. (2009). Polynomial calculation of the Shapley value based on sampling. *Computers & Operations Research*, *36*(5), 1726–1730.

Cheng, Y., Tan, J., & Lin, W.-H. (2013). Allocating the cost of empty railcar miles in a dynamic railroad network. *International Journal of Shipping and Transport Logistics*, *5*(3), 350–366.

Chinh, N. Q., Kim, H. C., Siwei, J., & NengSheng, Z. (2016). Collaborative vehicle routing problem for urban last-mile logistics. *2016 IEEE international conference on systems, man, and cybernetics (SMC)*, 001531–001536.

Christofides, N. (1979). The vehicle routing problem. *Combinatorial optimization*.

Clarke, G., & Wright, J. W. (1964). Scheduling of vehicles from a central depot to a number of delivery points. *Operations Research*, *12*(4), 568–581.

Cochran, W. G. (1977). *Sampling techniques*.

Cordeau, J.-F., Gendreau, M., & Laporte, G. (1997). A tabu search heuristic for periodic and multi-depot vehicle routing problems. *Networks: An International Journal*, *30*(2), 105–119.

Cordeau, J.-F., Laporte, G., Savelsbergh, M. W., & Vigo, D. (2007). Vehicle routing. *Handbooks in operations research and management science*, *14*, 367–428.

Cornuejols, G., & Harche, F. (1993). Polyhedral study of the capacitated vehicle routing problem. *Mathematical Programming*, *60*(1-3), 21–52.

Cruijssen, F., Borm, P., Fleuren, H., & Hamers, H. (2010). Supplier-initiated outsourcing: A methodology to exploit synergy in transportation. *European Journal of Operational Research*, *207*(2), 763–774.

Cruijssen, F., Bräysy, O., Dullaert, W., Fleuren, H., & Salomon, M. (2007). Joint route planning under varying market conditions. *International Journal of Physical Distribution & Logistics Management*, *37*(4), 287–304.

Cruijssen, F., Cools, M., & Dullaert, W. (2007). Horizontal cooperation in logistics: opportunities and impediments. *Transportation Research Part E: Logistics and Transportation Review*, *43*(2), 129–142.

Cruijssen, F., & Cruijssen, F. (2020). Applications of Cross-Chain Collaboration. *Cross-Chain Collaboration in Logistics: Looking Back and Ahead*, 91–105.

Cruijssen, F., Dullaert, W., & Fleuren, H. (2007). Horizontal cooperation in transport and logistics: a literature review. *Transportation journal*, *46*(3), 22–39.

Dahl, S., & Derigs, U. (2011). Cooperative planning in express carrier networks—An empirical study on the effectiveness of a real-time Decision Support System. *Decision Support Systems*, *51*(3), 620–626.

Dahle, L., Andersson, H., Christiansen, M., & Speranza, M. G. (2019). The pickup and delivery problem with time windows and occasional drivers. *Computers & Operations Research*, *109*, 122–133.

Dai, B., & Chen, H. (2011). A multi-agent and auction-based framework and approach for carrier collaboration. *Logistics Research*, *3*, 101–120.

Dai, B., & Chen, H. (2012). Profit allocation mechanisms for carrier collaboration in pickup and delivery service. *Computers & Industrial Engineering*, *62*(2), 633–643.

Dai, B., & Chen, H. (2015). Proportional egalitarian core solution for profit allocation games with an application to collaborative transportation planning. *European Journal of Industrial Engineering*, *9*(1), 53–76.

Dantzig, G. B., & Ramser, J. H. (1959). The truck dispatching problem. *Management science*, *6*(1), 80–91.

Defryn, C., Sörensen, K., & Cornelissens, T. (2016). The selective vehicle routing problem in a collaborative environment. *European Journal of Operational Research*, *250*(2), 400–411.

Defryn, C., Vanovermeire, C., & Sörensen, K. (2016). Gain sharing in horizontal logistic co-operation: a case study in the fresh fruit and vegetables sector. *Sustainable Logistics and Supply Chains: Innovations and Integral Approaches*, 75–89.

Deidda, D., Andreu, J., Perez, M., Sechi, G., Zucca, R., & Zuddas, P. (2009). A cooperative game theory approach to water pricing in a complex water resource system. *18th World IMACS Congress and MODSIM09 International Congress on Modelling and Simulation*, 3252–3258.

Deng, X., & Papadimitriou, C. H. (1994). On the complexity of cooperative solution concepts. *Mathematics of operations research*, *19*(2), 257–266.

Derks, J., & Kuipers, J. (1997). On the core of routing games. *International Journal of Game Theory*, *26*, 193–205.

Desrochers, M., & Laporte, G. (1991). Improvements and extensions to the Miller-Tucker-Zemlin subtour elimination constraints. *Operations Research Letters*, *10*(1), 27–36.

Dowdy, S., Wearden, S., & Chilko, D. (2011). *Statistics for research*. John Wiley & Sons.

Dror, M. (1990). Cost allocation: the traveling salesman, binpacking, and the knapsack. *Applied Mathematics and Computation*, *35*(2), 191–207.

Engevall, S., Göthe-Lundgren, M., & Värbrand, P. (1998). The traveling salesman game: An application ofcost allocation in a gas and oil company. *Annals of Operations Research*, *82*(0), 203–218.

Engevall, S., Göthe-Lundgren, M., & Värbrand, P. (2004). The heterogeneous vehicle-routing game. *Transportation Science*, *38*(1), 71–85.

Ergun, O., Kuyzu, G., & Savelsbergh, M. (2007). Reducing truckload transportation costs through collaboration. *Transportation science*, *41*(2), 206–221.

Estévez-Fernández, A., Borm, P., Meertens, M., & Reijnierse, H. (2009). On the core of routing games with revenues. *International Journal of Game Theory*, *38*, 291–304.

Faigle, U., Fekete, S. P., Hochstättler, W., & Kern, W. (1998). On approximately fair cost allocation in Euclidean TSP games. *Operations-Research-Spektrum*, *20*, 29–37.

Faigle, U., & Kern, W. (1992). The Shapley value for cooperative games under precedence constraints. *International Journal of Game Theory*, *21*, 249–266.

Fang, X., & Cho, S.-H. (2014). Stability and endogenous formation of inventory transshipment networks. *Operations Research*, *62*(6), 1316–1334.

Fernández, E., Roca-Riu, M., & Speranza, M. G. (2018). The shared customer collaboration vehicle routing problem. *European journal of operational research*, *265*(3), 1078–1093.

Ferrell, W., Ellis, K., Kaminsky, P., & Rainwater, C. (2020). Horizontal collaboration: opportunities for improved logistics planning. *International Journal of Production Research*, *58*(14), 4267–4284.

Fiestras-Janeiro, M. G., García-Jurado, I., Meca, A., & Mosquera, M. A. (2013). A new cost allocation rule for inventory transportation systems. *Operations Research Letters*, *41*(5), 449–453.

Fiestras-Janeiro, M. G., García-Jurado, I., Meca, A., & Mosquera, M. (2012). Cost allocation in inventory transportation systems. *TOP*, *20*, 397–410.

Fishburn, P., & Pollak, H. (1983). Fixed-route cost allocation. *The American Mathematical Monthly*, *90*(6), 366–378.

Flisberg, P., Frisk, M., Rönnqvist, M., & Guajardo, M. (2015). Potential savings and cost allocations for forest fuel transportation in Sweden: A country-wide study. *Energy*, *85*, 353–365.

Frisk, M., Göthe-Lundgren, M., Jörnsten, K., & Rönnqvist, M. (2010). Cost allocation in collaborative forest transportation. *European Journal of Operational Research*, *205*(2), 448–458.

Gansterer, M., & Hartl, R. F. (2018a). Centralized bundle generation in auction-based collaborative transportation. *Or Spectrum*, *40*(3), 613–635.

Gansterer, M., & Hartl, R. F. (2018b). Collaborative vehicle routing: A survey. *European Journal of Operational Research*, *268*(1), 1–12.

Gansterer, M., & Hartl, R. F. (2020). Shared resources in collaborative vehicle routing. *TOP*, *28*, 1–20.

Gansterer, M., Hartl, R. F., & Vetschera, R. (2019). The cost of incentive compatibility in auction-based mechanisms for carrier collaboration. *Networks*, *73*(4), 490–514.

Goel, A., & Gruhn, V. (2008). A general vehicle routing problem. *European Journal of Operational Research*, *191*(3), 650–660.

Göthe-Lundgren, M., Jörnsten, K., & Värbrand, P. (1996). On the nucleolus of the basic vehicle routing game. *Mathematical Programming*, *72*, 83–100.

Guajardo, M., Jörnsten, K., & Rönnqvist, M. (2016). Constructive and blocking power in collaborative transportation. *OR spectrum*, *38*, 25–50.

Guajardo, M., & Rönnqvist, M. (2015). Operations research models for coalition structure in collaborative logistics. *European Journal of Operational Research*, *240*(1), 147–159.

Guajardo, M., & Rönnqvist, M. (2016). A review on cost allocation methods in collaborative transportation. *International transactions in operational research*, *23*(3), 371–392.

Hamers, H., Borm, P., van de Leensel, R., & Tijs, S. (1999). Cost allocation in the Chinese postman problem. *European Journal of Operational Research*, *118*(1), 153–163.

Hansen, M. H., & Yu, B. (2001). Model selection and the principle of minimum description length. *Journal of the american statistical association*, 746–774.

Hezarkhani, B., Slikker, M., & Van Woensel, T. (2016). A competitive solution for cooperative truckload delivery. *Or Spectrum*, *38*, 51–80.

Hu, W., Dong, J., Hwang, B.-g., Ren, R., & Chen, Z. (2019). A scientometrics review on city logistics literature: Research trends, advanced theory and practice. *Sustainability*, *11*(10), 2724.

Iannaccone, G., Marcucci, E., & Gatta, V. (2021). What young e-consumers want? Forecasting parcel lockers choice in Rome. *Logistics*, *5*(3), 57.

Irnich, S., Toth, P., & Vigo, D. (2014). Chapter 1: The family of vehicle routing problems. In *Vehicle Routing: Problems, Methods, and Applications, Second Edition* (pp. 1–33). SIAM.

Jílková, P., & Králová, P. (2021). Digital consumer behaviour and ecommerce trends during the COVID-19 crisis. *International Advances in Economic Research*, *27*(1), 83–85.

Kallehauge, B., Larsen, J., Madsen, O. B., & Solomon, M. M. (2005). *Vehicle routing problem with time windows*. Springer.

Karaenke, P., Bichler, M., & Minner, S. (2019). Coordination is hard: Electronic auction mechanisms for increased efficiency in transportation logistics. *Management Science*, *65*(12), 5884–5900.

Krajewska, M. A., & Kopfer, H. (2006). Collaborating freight forwarding enterprises: Request allocation and profit sharing. *OR spectrum*, *28*, 301–317.

Krajewska, M. A., Kopfer, H., Laporte, G., Ropke, S., & Zaccour, G. (2008). Horizontal cooperation among freight carriers: request allocation and profit sharing. *Journal of the Operational Research Society*, *59*, 1483–1491.

Kritikos, M. N., & Ioannou, G. (2010). The balanced cargo vehicle routing problem with time windows. *International Journal of Production Economics*, *123*(1), 42–51.

Laporte, G. (1992). The traveling salesman problem: An overview of exact and approximate algorithms. *European Journal of Operational Research*, *59*(2), 231–247.

Laporte, G. (2009). Fifty years of vehicle routing. *Transportation science*, *43*(4), 408–416.

Lehoux, N., d'Amours, S., Frein, Y., Langevin, A., & Penz, B. (2011). Collaboration for a two-echelon supply chain in the pulp and paper industry: the use of incentives to increase profit. *Journal of the Operational Research Society*, *62*(4), 581–592.

Leng, K., Ya, B., Linbo, J., Han-Chi, F., & Van Nieuwenhuyse, I. (2018). Research on agricultural supply chain system with double chain architecture based on blockchain technology. *Future Generation Computer Systems*, *86*(1), 641–649.

Littlechild, S. C., & Owen, G. (1973). A simple expression for the Shapley value in a special case. *Management Science*, *20*(3), 370–372.

Liu, P., Wu, Y., & Xu, N. (2010). Allocating collaborative profit in less-than-truckload carrier alliance. *Journal of Service science and Management*, *3*(01), 143.

Lozano, S., Moreno, P., Adenso-Díaz, B., & Algaba, E. (2013). Cooperative game theory approach to allocating benefits of horizontal cooperation. *European Journal of Operational Research*, *229*(2), 444–452.

Massol, O., & Tchung-Ming, S. (2010). Cooperation among liquefied natural gas suppliers: Is rationalization the sole objective? *Energy Economics*, *32*(4), 933–947.

Miller, C. E., Tucker, A. W., & Zemlin, R. A. (1960). Integer programming formulation of traveling salesman problems. *Journal of the ACM (JACM)*, *7*(4), 326–329.

Moretti, S., & Patrone, F. (2008). Transversality of the Shapley value. *TOP*, *16*, 1–41.

Moretti, S., Patrone, F., & Bonassi, S. (2007). The class of microarray games and the relevance index for genes. *TOP*, *15*, 256–280.

Moulin, H. (1991). *Axioms of cooperative decision making*. Cambridge university press.
Muñoz-Villamizar, A., Montoya-Torres, J. R., & Vega-Mejía, C. A. (2015). Non-collaborative versus collaborative last-mile delivery in urban systems with stochastic demands. *Procedia CIRP*, *30*, 263–268.
Nguyen, C., Dessouky, M., & Toriello, A. (2014). Consolidation strategies for the delivery of perishable products. *Transportation Research Part E: Logistics and Transportation Review*, *69*, 108–121.
Osborne, J. W., & Waters, E. (2002). Four assumptions of multiple regression that researchers should always test. *Practical assessment, research, and evaluation*, *8*(1), 2.
Özener, O. Ö. (2014). Developing a collaborative planning framework for sustainable transportation. *Mathematical Problems in Engineering*, *2014*.
Özener, O. Ö., & Ergun, Ö. (2008). Allocating costs in a collaborative transportation procurement network. *Transportation Science*, *42*(2), 146–165.
Özener, O. Ö., Ergun, Ö., & Savelsbergh, M. (2011). Lane-exchange mechanisms for truckload carrier collaboration. *Transportation Science*, *45*(1), 1–17.
Özener, O. Ö., Ergun, Ö., & Savelsbergh, M. (2013). Allocating cost of service to customers in inventory routing. *Operations Research*, *61*(1), 112–125.
Pérez-Bernabeu, E., Juan, A. A., Faulin, J., & Barrios, B. B. (2015). Horizontal cooperation in road transportation: a case illustrating savings in distances and greenhouse gas emissions. *International Transactions in Operational Research*, *22*(3), 585–606.
Potters, J. A., Curiel, I. J., & Tijs, S. H. (1992). Traveling salesman games. *Mathematical Programming*, *53*, 199–211.
Prins, C. (2004). A simple and effective evolutionary algorithm for the vehicle routing problem. *Computers & operations research*, *31*(12), 1985–2002.
Puka, R., Duda, J., & Karkula, M. (2019). Improving of Selected Routing Strategies for Order Picking. *International Scientific Conference Transport of the 21st Century*, 363–373.
Ranieri, L., Digiesi, S., Silvestri, B., & Roccotelli, M. (2018). A review of last mile logistics innovations in an externalities cost reduction vision. *Sustainability*, *10*(3), 782.
Rawlings, J. O., Pantula, S. G., & Dickey, D. A. (1998). *Applied regression analysis: a research tool*. Springer.
Sanchez, M., Pradenas, L., Deschamps, J.-C., & Parada, V. (2016). Reducing the carbon footprint in a vehicle routing problem by pooling resources from different companies. *NETNOMICS: Economic Research and Electronic Networking*, *17*, 29–45.
Schmeidler, D. (1969). The nucleolus of a characteristic function game. *SIAM Journal on applied mathematics*, *17*(6), 1163–1170.
Schmelzer, H., Bütikofer, S., & Hollenstein, L. (2016). Kooperieren? ja! aber wie?: Chancen und Herausforderungen bei der Entwicklung einer Kooperationsplattform für urbane Güterlogistik in der Stadt Zürich. *Logistics Innovation*, *2016*(1), 16–19.
Schulte, F., Lalla-Ruiz, E., González-Ramírez, R. G., & Voß, S. (2017). Reducing port-related empty truck emissions: a mathematical approach for truck appointments with collaboration. *Transportation Research Part E: Logistics and Transportation Review*, *105*, 195–212.
Shapley, L. S. (1951). Notes on the n-person game. *Rand Corporation*.
Shapley, L. S. (1953). A value for n-person games. *Annals of Math. Studies*, *28*.
Shi, Y., Lin, N., Han, Q., Zhang, T., & Shen, W. (2020). A method for transportation planning and profit sharing in collaborative multi-carrier vehicle routing. *Mathematics*, *8*(10), 1788.
Solomon, M. M. (1987). Algorithms for the vehicle routing and scheduling problems with time window constraints. *Operations Research*, *35*(2), 254–265.
Song, L., Cherrett, T., McLeod, F., & Guan, W. (2009). Addressing the last mile problem: transport impacts of collection and delivery points. *Transportation research record*, *2097*(1), 9–18.
Sun, L., Rangarajan, A., Karwan, M. H., & Pinto, J. M. (2015). Transportation cost allocation on a fixed route. *Computers & Industrial Engineering*, *83*, 61–73.
Sykes, A. O. (1993). An introduction to regression analysis.
Tang, Q., Zhang, Z., Yuan, Z., & Li, Z. (2022). The Game Analysis of Information Sharing for Supply Chain Enterprises in the Blockchain. *Frontiers in Manufacturing Technology*, *2*, 880332.
Tarantilis, C. D. (2005). Solving the vehicle routing problem with adaptive memory programming methodology. *Computers & Operations Research*, *32*(9), 2309–2327.
Toriello, A., & Uhan, N. A. (2013). On traveling salesman games with asymmetric costs. *Operations Research*, *61*(6), 1429–1434.
Toth, P., & Vigo, D. (2014). *Vehicle routing: problems, methods, and applications*. SIAM.
van Campen, T., Hamers, H., Husslage, B., & Lindelauf, R. (2018). A new approximation method for the Shapley value applied to the WTC 9/11 terrorist attack. *Social Network Analysis and Mining*, *8*, 1–12.
Vanovermeire, C., & Sörensen, K. (2014a). Integration of the cost allocation in the optimization of collaborative bundling. *Transportation Research Part E: Logistics and Transportation Review*, *72*, 125–143.
Vanovermeire, C., & Sörensen, K. (2014b). Measuring and rewarding flexibility in collaborative distribution, including two-partner coalitions. *European Journal of Operational Research*, *239*(1), 157–165.
Vanovermeire, C., Sörensen, K., Van Breedam, A., Vannieuwenhuyse, B., & Verstrepen, S. (2014). Horizontal logistics collaboration: decreasing costs through flexibility and an adequate cost allocation strategy. *International Journal of Logistics Research and Applications*, *17*(4), 339–355.
VROOM Project. (2022). Vehicle Routing Open Source Optimization Machine. http://vroom-project.org/
Wang, Y., Ma, X., Xu, M., Wang, L., Wang, Y., & Liu, Y. (2015). A methodology to exploit profit allocation in logistics joint distribution network optimization. *Mathematical Problems in Engineering*, *2015*.
Wong, H., Van Oudheusden, D., & Cattrysse, D. (2007). Cost allocation in spare parts inventory pooling. *Transportation Research Part E: Logistics and Transportation Review*, *43*(4), 370–386.
Yang, F., Yang, M., Xia, Q., & Liang, L. (2016). Collaborative distribution between two logistics service providers. *International Transactions in Operational Research*, *23*(6), 1025–1050.
Yengin, D. (2012). Characterizing the Shapley value in fixed-route traveling salesman problems with appointments. *International Journal of Game Theory*, *41*, 271–299.
Yilmaz, O., & Savasaneril, S. (2012). Collaboration among small shippers in a transportation market. *European journal of operational research*, *218*(2), 408–415.
Yule, G. U., & Kendall, M. (1968). *An introduction to the theory of Statistics*. Charles Griffin; Co. Ltd.
Zakharov, V. V., & Shchegryaev, A. N. (2015). Stable cooperation in dynamic vehicle routing problems. *Automation and Remote Control*, *76*, 935–943.